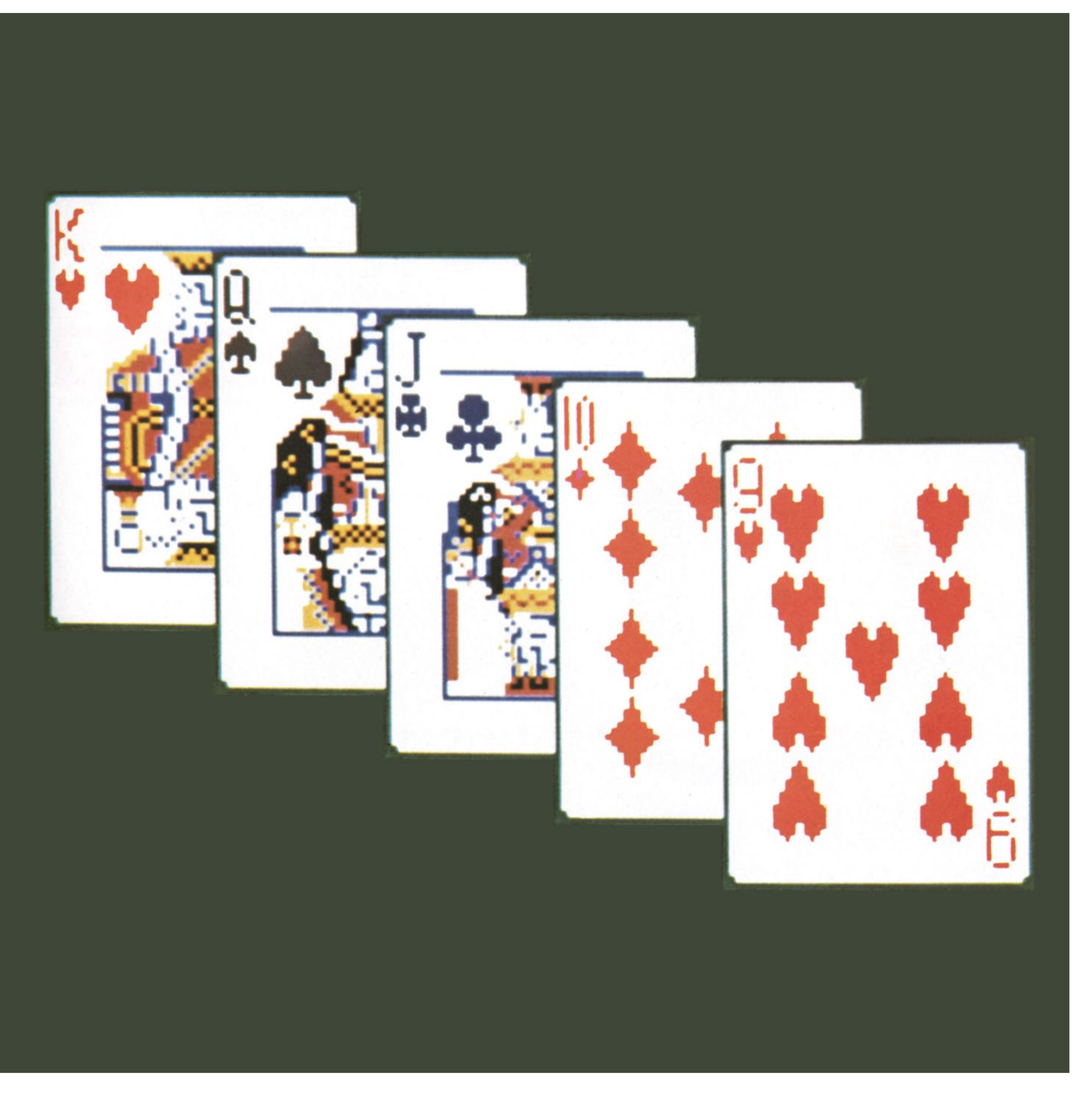
INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN



POKER POR ORDENADOR

Septiembre 1978

170 Ptas.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 24
Septiembre 1978

INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN

EDITA Prensa Científica, S.A.
Calabria, 235-239
Barcelona-29 (ESPAÑA)

Colaboradores de este número:

Asesoramiento y traducción:

Monique Robert: *La diferenciación química de las células nerviosas;* Manuel Puigcerver: *Los rayos cósmicos de máxima energía;* Amando García: *Imágenes nocturnas de la Tierra vista desde el espacio;* Oleguer Escolá: *Aprendizaje y memoria de las abejas;* Lluïsa Vilageliu: *La tecnología de las membranas sintéticas;* Joandomènec Ros: *El kiwi;* Luis Bou: *Póker computerizado; Juegos matemáticos;* José E. Myro y Alicia Menéndez: *Taller y laboratorio*

Ciencia y sociedad:

Jorge Domingo

Libros:

Javier de Lorenzo y José Simón Díaz.

LA PORTADA

La fotografía de la portada muestra una mano de póker (una escalera al rey) que aparece en la pantalla del tubo de rayos catódicos de un terminal de computador. Hoy día se están programando ordenadores para jugar al póker, y dicha terminal puede usarse para controlar juegos en los que alguno o todos los jugadores sean programas de ordenador (ver "Póker computerizado", por Nicholas V. Findler, en este número). El póker es un juego de especial interés porque sirve de modelo para muchas situaciones de toma de decisiones en la vida real que implican riesgo e incertidumbre, verbigracia: planificación económica, campañas políticas y estrategia militar. Al crear estos programas, los científicos esperan llegar a una mejor comprensión del mecanismo de toma de decisiones y, así, aprender a utilizar los ordenadores para resolver problemas que siempre se reputaron ser competencia exclusiva del juicio humano.


Suscripciones:

Prensa Científica, S.A.
Departamento de suscripciones
Calabria 235-239
Barcelona-29 (España)
Teléfono 322 05 51 ext. 41

Distribución para España:

Distribuciones de Enlace, S.A.
Ausias March, 49
Barcelona-10 (España)

Distribución para los restantes países:

Editorial Labor, S.A.
Calabria, 235-239
Barcelona-29 (España)

Precio de venta de este número:

España: 170 pesetas
Extranjero: 2 U.S.$

Condiciones de suscripción:

España:
Un año (12 números): 1870 pesetas
Extranjero:
Un año (12 números): 30 U.S.$

Ejemplar atrasado ordinario:
190 pesetas

Ejemplar atrasado extraordinario:
270 pesetas

Difusión controlada por **OJD**

PROCEDENCIA DE LAS ILUSTRACIONES

Fotografía de la portada de Ralph Morse

Página	Fuente	Página	Fuente
5-12	P. Pascual y R. Tarrach	47	© 1969 National Geographic Society *(arriba);* NASA *(abajo)*
15	Dennis Bray, Medical Research Council Cell Biophysics Unit	48	*Thomas A. Croft, SRI International (arriba),* Allen Beechel *(abajo)*
16	Bunji Tagawa	49	Allen Beechel
17	Paul H. Patterson, Harvard Medical School	50-51	U.S. Air Force
18	Bunji Tagawa	52	U.S. Air Force *(arriba),* Bureau of the Census *(abajo)*
19	Paul H. Patterson, Harvard Medical School	53-54	U.S. Air Force
20	Bunji Tagawa	57-62	Patricia J. Wynne
21	Story C. Landis, Harvard Medical School *(arriba y abajo, izquierda);* Bunji Tagawa *(abajo, derecha)*	63	Randolf Menzel, Universidad Libre de Berlín
		64	Patricia J. Wynne
		67	Amicon Corporation
		68-79	George V. Kelvin
22-23	Bunji Tagawa	83-85	Alan D. Iselin
27	Ilil Arbel	86	Cynthia Carey, Universidad de Colorado
28	Meridian Airmaps Ltd.	87-90	Alan D. Iselin
29-34	Ilil Arbel	93	Ralph Morse
35	Eugene Loh, Universidad de Utah *(arriba);* Ilil Arbel *(abajo)*	94-98	Jerome Kuhl
		103-106	Ilil Arbel
		109	Fritz Goro
		110	Michael Goodman
36	Ilil Arbel	111	Michael Goodman *(arriba),* Fritz Goro *(abajo)*
38-39	Jorge Domingo		
43-46	U.S. Air Force	112	Michael Goodman

Dep. legal: B. 38.999-76
Fotocomposición: Tecfa
Guipúzcoa, 36 (local 1) Barcelona-20
Cayfosa. Santa Perpetua de Moguda
Barcelona.
Printed in Spain - Impreso en España

Los autores

PEDRO PASCUAL Y ROLF TARRACH ("Monopolos") trabajan en el Departamento de Física Teórica de la Universidad de Barcelona. El primero ha sido miembro de la Junta de Energía Nuclear y catedrático de la Universidad de Valencia, cargo que ocupa hoy en la de Barcelona. Pasó dos años en la Universidad de Chicago como becario de la International Atomic Energy Agency (IAEA). Posteriormente ha venido realizando varias estancias en el CERN contratado por dicho organismo europeo como "Fellow Correspondant" y "Visiting Scientist". Ha dirigido el Grupo Interuniversitario de Física Teórica (GIFT) y pertenece hoy al European Committee for Future Accelerator. El profesor Tarrach, quien también estuvo en la Universidad de Valencia, ocupa el cargo de profesor agregado interino en la Universidad de Barcelona. Ha pasado asimismo por la Universidad de Chicago y, por el CERN, en calidad de becario del GIFT primero y luego contratado como "Visiting Scientist"

PAUL H. PATTERSON, DAVID D. POTTER Y EDWIN J. FURSHPAN ("La diferenciación química de las células nerviosas") son neurobiólogos adscritos a la Facultad de Medicina de Harvard. Patterson se halla en Harvard desde hace ocho años; se doctoró en bioquímica en la Universidad Johns Hopkins. Potter y Furshpan llegaron a Harvard en 1959, después de colaborar como becarios postdoctorales en el departamento de biofísica del University College London. Potter se doctoró en biología por Harvard y Furshpan en fisiología animal por el Instituto de Tecnología de California. En conjunto, los trabajos descritos en este artículo fueron realizados con la colaboración de Dennis Bray, Robert B. Campenot, Linda L. Y. Chun, Philippa Claude, Karen Fisher, Story C. Landis, Peter R. MacLeish, Doreen McDowell, Richard E. Mains, Colin A. Nurse, Kunihiko Obata, Paul H, O'Lague, Louis F. Reichardt, Patricia A. Walicke y Michel Weber.

JOHN LINSLEY ("Los rayos cósmicos de máxima energía") es profesor de investigación de física en la Universidad de Nuevo México. Obtuvo sus títulos en la Universidad de Minnesota, donde participó en experimentos proyectados para estudiar la composición de los rayos cósmicos primarios mediante una variedad de técnicas todas las cuales incluían el uso de globos para grandes alturas. En 1954 entró a formar parte del grupo del Instituto de Tecnología de Massachusetts (MIT) dirigido por Bruno Rossi, para trabajar en el sistema de detección de chubascos aéreos de la Estación Agassiz, primera de tales instalaciones capaz de determinar tanto la energía como la dirección de los chubascos aéreos de rayos cósmicos. Cuatro años más tarde, él y Livio Scarsi montaron el primer dispositivo gigante de detección de rayos cósmicos, concebido según la instalación de Agassiz, en Rancho Volcano, cerca de Albuquerque, Nuevo México.

THOMAS A. CROFT ("Imágenes nocturnas de la Tierra vista desde el espacio") trabaja en el SRI Internacional (antes Stanford Research Institute). Después de graduarse en el Dartmouth College, Croft fue piloto de caza de la Armada y más tarde trabajó como ingeniero aerospacial. Esta última actividad determinó su inclinación a la investigación; en consecuencia, regresó a la vida académica, y en 1965 obtuvo el doctorado en la Universidad de Stanford. Continuó trabajando en Stanford hasta 1976, fecha en que pasó al puesto que ocupa en la actualidad. En un principio, su trabajo de investigación se centró en el campo de la transmisión de radio en la ionosfera y, en particular, sobre simulación con computadores digitales. Croft ha realizado también ciertas experiencias sobre el viento solar. Fue miembro de los equipos científicos encargados del Pioneer y del Voyager.

RANDOLF MENZEL Y JOCHEN ERBER ("Aprendizaje y memoria de las abejas") trabajan en el Instituto de Fisiología Animal de la Universidad Libre de Berlín, en donde Menzel es catedrático y Erber profesor ayudante. Ambos llegaron a Berlín en 1976 después de haber investigado durante un año en la Universidad Nacional Australiana de Canberra. Con anterioridad se habían encontrado en la Universidad Técnica de Darmstadt: Menzel era allí profesor ayudante y Erber alumno graduado. Menzel había estudiado zoología, química y física en la Universidad de Francfurt y en la de Tübingen; defendió la tesis doctoral en zoología en Tübingen en 1967. Erber, que empezó estudiando ingeniería eléctrica en Darmstadt, también leyó su tesis doctoral en zoología en dicho centro.

NICHOLAS V. FINDLER ("Póker computerizado") es profesor de ciencias de la computación en la Universidad estatal de Nueva York en Buffalo. Es titulado en ingeniería eléctrica y doctor en física matemática por la Universidad de Ciencias Técnicas de Budapest. Abandonó Hungría en 1956, y tras una breve permanencia en Viena, fue a vivir a Australia, donde ha trabajado hasta 1963, en que se trasladó a los Estados Unidos. Ha venido ocupando su puesto actual desde 1966, excepto durante una licencia sabática en 1972-73, en que fue Senior Fulbright Scholar en la Universidad Técnica de Viena.

HARRY P. GREGOR Y CHARLES D. GREGOR ("La tecnología de las membranas sintéticas") son padre e hijo. Gregor padre es catedrático de ingeniería química y química aplicada en la Universidad de Columbia. Realizó su tesis doctoral en la Universidad de Minnesota, donde estableció contacto con científicos refugiados alemanes, en particular con Karl Sollner, bajo cuya dirección fabricó las primeras membranas "permselectivas". Desde entonces ha trabajado en la síntesis, caracterización y utilización de varios tipos de membranas; posee numerosas patentes de membranas y tratamientos con membranas. Su hijo Charles se graduó en el Carleton Collage y es doctor por la Universidad de Columbia. Estuvo el año pasado en Jerusalén estudiando el Talmud, y ahora está en la Universidad de New York en Stony Brook, donde investiga en el campo de la química fisico-orgánica.

WILLIAM A. CALDER III ("El kiwi") es profesor de ecología y de biología evolutiva en la Universidad de Arizona. Se graduó por la Universidad de Georgia y obtuvo su doctorado en la Universidad Duke en 1966; enseñó en Duke y en el Instituto Politécnico de Virginia antes de trasladarse a Arizona en 1969. Además de sus tareas universitarias de docencia e investigación, es socio vitalicio y miembro del Rocky Mountain Biological Laboratory. En 1976-1977, subvencionado por la Nacional Science Foundation, escribe, "pasé ese año sabático dedicado sobre todo a los curiosos huevos, a la incubación y el metabolismo de los kiwis de Nueva Zelanda, aves únicas entre las aves en todos los aspectos, pero que los americanos sólo conocen a través de la marca de fábrica de latas de betún para los zapatos."

Monopolos

La existencia de monopolos, que son cargas magnéticas libres, permitiría dar mayor simetría a las ecuaciones fundamentales del electromagnetismo y explicar además la cuantificación observada de la carga eléctrica

Pedro Pascual y Rolf Tarrach

La electricidad y el magnetismo tienen su origen en tiempos muy remotos. En la Grecia antigua, hace más de 2000 años, se conocían dos hechos básicos de estas ramas de la física: que si se frotaba un trozo de ámbar (en griego ἤλεκτρον) con un paño se lograba hacer mover objetos muy ligeros en virtud de lo que ahora se llaman fuerzas electrostáticas y que ciertas piedras procedentes de Magnesia, ahora denominadas magnetitas, atraían trocitos de hierro. Para los griegos éstos eran fenómenos sin ningún interés práctico. Habrá que esperar hasta el s. XII para encontrar referencias sobre el uso del imán en la navegación.

En el año 1600 aparece la obra fundamental de William Gilbert *De magnete magneticisque corporibus et de magno magnete tellura physiologia nova*, en la que desarrolla, dentro del espíritu de la ciencia escolástica, extensas teorías de los fenómenos electrostáticos y magnéticos; en ella se da, por primera vez, la razón de que la aguja de la brújula se oriente por sí misma en una dirección definida: la Tierra es un imán permanente. Se puede considerar esta obra como el hito que marca el nacimiento de la electricidad y el magnetismo como ciencias. A pesar de la obra de Gilbert, el avance de estas dos ciencias, que juegan un papel fundamental en la vida moderna, fue extraordinariamente lento en los 200 años siguientes, centrándose muchos de los estudios en el magnetismo en razón de su importancia en la navegación. Sin embargo, merecen mencionarse, entre otros, el descubrimiento de la máquina electrostática por Otto von Guericke en 1672, a cuya mejora se dedicaron un gran número de científicos y técnicos a lo largo de todo el siglo XVIII; el condensador, descubierto independientemente en 1745 por Ewald Georg von Kleist y Petrus van Musschenbroek, así como los estudios de la electricidad atmosférica iniciados a mediados del siglo XVIII por el norteamericano Benjamin Franklin, los cuales despertaron gran interés y numerosas controversias sobre si los hombres debían interferir, con este tipo de experiencias, en los fenómenos naturales considerados en general como la expresión directa de la voluntad de Dios.

A finales del siglo XVIII se dieron dos pasos realmente significativos. Por un lado, Charles Augustin Coulomb, en 1785, formuló matemáticamente la ley fundamental de toda la electrostática, que dice: La fuerza ejercida por una carga puntual sobre otra está dirigida a lo largo de la línea que las une. Es repulsiva si las cargas tienen el mismo signo y atractiva si las cargas tienen signos opuestos. La fuerza varía inversamente con el cuadrado de la distancia y es proporcional al valor de cada una de las cargas. Esta ley se puede resumir en una simple ecuación de la forma siguiente: Sea **r** el vector que va desde la posición de la carga q_1 a la de la carg q_2, cuyo módulo (esto es, la distancia entre las dos cargas) indicaremos por r, entonces la ley de Coulomb dice que la fuerza (**F**) ejercida por la partícula de carga q_1 sobre la de carga q_2 es $\mathbf{F} = (q_1 q_2 / r^3)\mathbf{r}$. Si las dos cargas tienen el mismo signo, el producto $q_1 q_2$ es positivo y **F** tiene la misma dirección y sentido que **r**, y, por tanto, las cargas se repelen; por el contrario, si las cargas tienen signos opuestos, el producto $q_1 q_2$ es negativo y **F** tiene la misma dirección y sentido opuesto a **r** y las cargas se atraen. La intensidad de la fuerza es el módulo de **F** cuyo valor es $|q_1 q_2|/r$.

Consideremos una serie de cargas eléctricas distribuidas en el espacio. En un punto arbitrario P coloquemos una carga pequeña q. Debido a la presencia de las otras cargas, una fuerza $\mathbf{F}(\mathbf{r},t)$ actuará sobre la carga situada en el punto P, siendo **r** las coordenadas de posición y t el tiempo. Esa fuerza será precisamente la suma de las fuerzas ejercidas por cada una de las cargas sobre q. Al escribir $\mathbf{F}(\mathbf{r},t)$ queremos indicar que esta fuerza dependerá tanto del punto del espacio en que se halle situada la carga q como del instante de tiempo que se considere, pues las cargas que originan esta fuerza pueden cambiar su posición de un instante a otro. Es extraordinariamente útil asociar al sistema de cargas considerado un campo eléctrico $\mathbf{E}(\mathbf{r},t)$, cuyo valor en cada punto del espacio y en cada instante de tiempo se obtiene dividiendo $\mathbf{F}(\mathbf{r},t)$ por la carga q, y representa la fuerza que actuaría sobre la carga eléctrica unidad positiva, situada en reposo en el punto de coordenadas **r**, en el instante t. La presencia de un campo eléctrico en una región del espacio puede indicarse, para un instante de tiempo determinado, de una forma intuitiva dibujando las llamadas líneas de fuerza; éstas se caracterizan por el hecho de que su tangente en cada punto coincide en dirección y sentido con las del campo eléctrico en dicho punto. Sucede además que dos líneas de fuerza no pueden cortarse en ningún punto del espacio, pues si tal sucediera, en este punto el campo eléctrico, al estar dirigido según la tangente a la línea de fuerza, no tendría una dirección bien definida, contra la misma definición de campo eléctrico. En la ilustración de la página siguiente se dibujan los esquemas de líneas de fuerza para algunas situaciones sencillas.

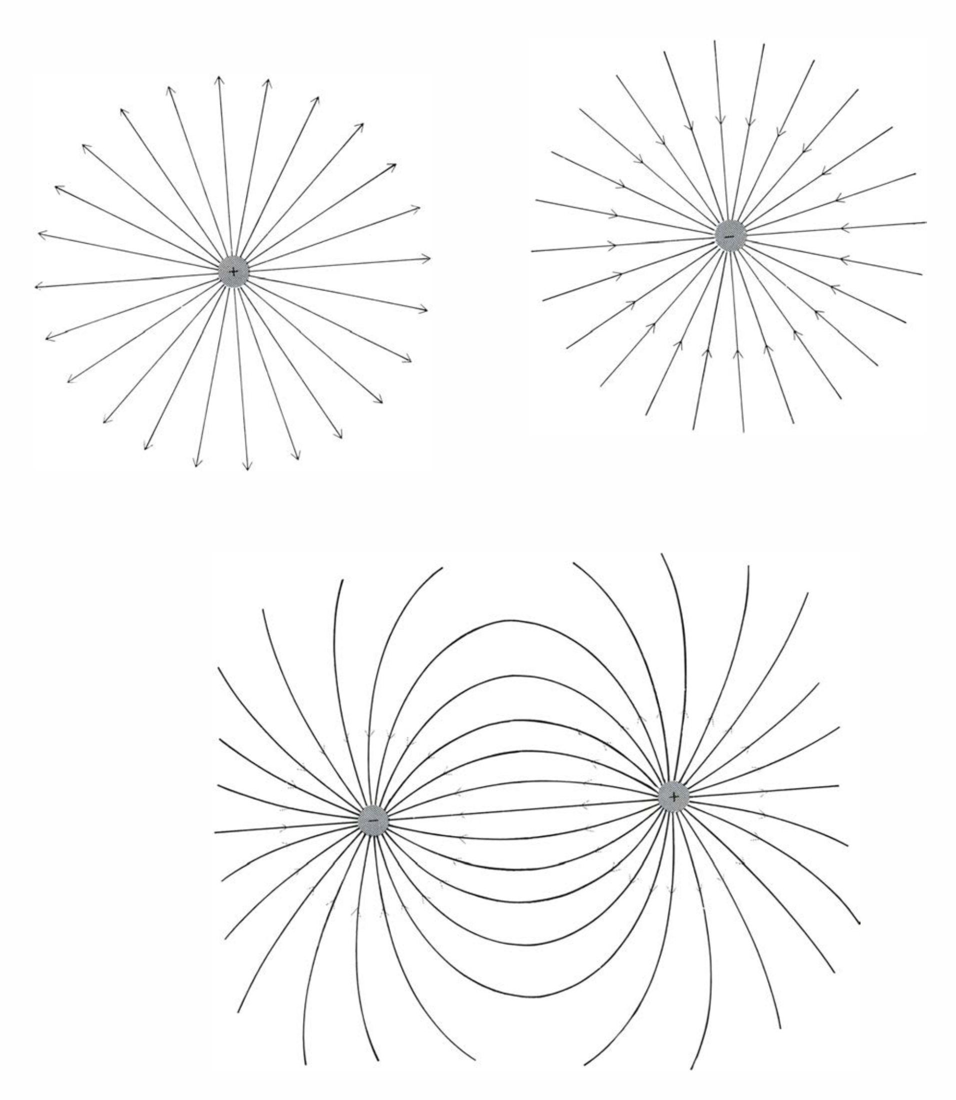
Otro importante concepto que debemos introducir es el de flujo de campo eléctrico a través de una superficie. Si en una región del espacio hay un campo eléctrico uniforme **E** y en ella se considera una superficie plana de área A cuyo vector normal forma un ángulo θ con la dirección del campo entonces el flujo viene dado por $\Phi = EA \cos\theta$. Nótese que el flujo puede ser positivo o negativo según que el ángulo que el campo eléctrico forme con la normal, que define la orientación de la superficie, sea menor o mayor que un ángulo recto. Si el campo eléctrico no es uniforme y la superficie tiene una forma cualquiera se puede calcular el flujo descomponiendo la superficie en superficies muy pequeñas, orientadas de forma continua, cada una de las cuales pueda considerarse como plana y tales que en la vecindad de cada una de ellas el campo eléctrico pueda considerarse como uniforme; se calcula el flujo a través de cada una de las superficies elementales

y el flujo total es la suma algebraica de los flujos así calculados. Teniendo en cuenta la interpretación de las líneas de fuerza resulta que el flujo a través de una superficie, una vez fijada de forma continua la normal en cada punto de la misma, es proporcional al número que se obtiene contando el número de líneas de fuerza que atraviesan la superficie considerada, tomando como positivas las que forman con la normal un ángulo menor que un recto y como negativas las que lo forman mayor que un recto. Carl Friedrich Gauss es el autor de un teorema muy útil relacionado con el flujo: consideremos una superficie cerrada arbitraria; de acuerdo con el procedimiento usual supondremos que la normal en cada punto de la superficie está dirigida hacia el exterior; entonces, el flujo del campo eléctrico a través de dicha superficie es igual a $4\pi Q$, donde Q es la suma algebraica de todas las cargas interiores a dicha superficie. En representaciones gráficas tales como las de la ilustración de esta página esto se traduce, como se comprueba fácilmente, en que si dibujamos una línea cerrada arbitraria el número de líneas de fuerza que la atraviesan es proporcional a la carga en el interior de la misma.

El otro descubrimento importante de finales del siglo XVIII fue el de la corriente eléctrica, como resultado de los trabajos del anatomista Aloisio Galvani en los años 1780-90, quien la presentó en 1791 en un libro titulado *De viribus electricitatis in motu musculari commentarius*. (Sin embargo, Alessandro Volta reclama el honor de haber aislado el fenómeno físico del contexto fisiológico en que lo describiera Galvani; Volta fue también el creador de lo que hoy día se conoce como pila voltaica.) Sea una cierta región del espacio en la que existe una distribución de cargas eléctricas cuya densidad indicaremos por $\varrho_e(\mathbf{r},t)$ para señalar que puede depender del punto del espacio y del instante de tiempo considerados; estas cargas pueden ser electrones de conducción en el interior de un metal, protones en el interior del tubo de un acelerador, iones con carga positiva o negativa en un electrolito, etc. A partir de una densidad podremos calcular la carga contenida en un determinado volumen del espacio y en un cierto instante, de la siguiente forma: dividiremos este volumen en pequeños elementos de modo que en cada uno de ellos la densidad de carga pueda considerarse constante; la carga contenida en este elemento de volumen es simplemente el producto de su volumen por la densidad de carga en él; la suma algebraica de las cargas así calculadas nos da la carga contenida en el elemento de volumen considerado.

Supongamos ahora que sobre esta región establecemos un campo eléctrico externo, las cargas positivas tienden a moverse en el sentido del campo eléctrico aplicado y las negativas en sentido contrario a él, creándose una corriente eléctrica. Esta corriente se caracterizará por un vector $\mathbf{j}_e(\mathbf{r},t)$, llamado densidad de corriente eléctrica, el flujo de esta corriente a través de una superficie nos da la carga neta que atraviesa ésta en la unidad de tiempo. Por ejemplo, si consideramos una distribución homogénea de partículas, y es n el número de éstas por unidad de volumen, todas ellas con una misma carga q y bajo la acción del campo eléctrico externo se mueven con una velocidad $\mathbf{v}$, entonces la densidad de corriente viene dada por $\mathbf{j}_e = nq\mathbf{v}$. Nótese que se toma como sentido de la corriente el correspondiente al movimiento de las cargas positivas. La densidad de carga y la densidad de corriente eléctrica no son dos cantidades totalmente arbitrarias sino que deben satisfacer la llamada ecuación de continuidad, que refleja una de las leyes fundamentales de la Física: la ley de conservación de la carga eléctrica. Esta ecuación de continuidad refleja matemáticamente el hecho de que si consideramos una superficie cerrada que limita un volumen del espacio, a partir de la densidad de carga se puede calcular en todo instante la carga en el interior de este volumen; las variaciones que se encuentren deben ser forzosamente debidas al flujo de cargas a través de la superficie que limita dicho volumen, el cual puede calcularse a partir de la densidad de corriente eléctrica.

Todo imán tiene un polo Norte o posi-



LINEAS DE FUERZA para algunos casos sencillos. Puesto que existen un número infinito de puntos, sólo se dibujan algunas líneas representativas y se indica el campo dibujando líneas de fuerza continuas que empiezan en las cargas positivas y terminan en las negativas. Puede ocurrir también que empiecen o terminen a distancia infinita. Para hacer estos dibujos se adopta el criterio de que el número de líneas de fuerza que salen de una carga puntual positiva o las que llegan a una carga puntual negativa sea proporcional al valor de esta carga; además las líneas de fuerza se dibujan simétricamente saliendo o entrando en la carga puntual. Se puede entonces comprobar que debido a la validez de la ley de Coulomb la densidad de líneas, es decir el número de ellas por unidad de área perpendicular a las mismas, es proporcional a la intensidad del campo eléctrico. En el extremo superior izquierdo, líneas de fuerza alrededor de una carga puntual $+q$, en el superior derecho las correspondientes a una carga $-q$. Abajo, líneas de fuerza creadas por un dipolo eléctrico (un sistema formado por una carga $+q$ y otra $-q$ separadas por una distancia L.)

tivo y un polo Sur o negativo de igual intensidad. Al dividir un imán aparecen dos polos iguales y opuestos en la zona de rotura, de modo que se obtienen dos imanes, ambos con polos de igual intensidad y de carácter opuesto. No se conoce ningún método para obtener cargas magnéticas aisladas, si es que éstas existen como tales. Un imán interacciona con otro imán de forma análoga a como lo hacen dos dipolos eléctricos; para describir el efecto producido por uno o más imanes sobre otro imán se introduce el concepto de campo magnético $\mathbf{B}(\mathbf{r},t)$ que se puede describir gráficamente introduciendo líneas de fuerza magnéticas, en total analogía a lo realizado para los campos eléctricos. Así por ejemplo, las líneas de fuerza asociadas al campo magnético creado por un imán en forma de barra son totalmente idénticas en su exterior a las líneas de fuerza eléctricas creadas por un dipolo, si ponemos el imán de forma que su polo positivo y su polo negativo coincidan, respectivamente, con la posición de la carga eléctrica positiva y negativa del dipolo.

Hace ya mucho tiempo que se sospechaba que la electricidad y el magnetismo, desarrollados paralela e independientemente, debían estar íntimamente relacionados. El año 1820 es una fecha crucial para la historia de estas ciencias, pues fue entonces cuando Hans Christian Oersted demostró experimentalmente que existía una relación clara entre la electricidad y el magnetismo. Las experiencias de Oersted fueron el primer paso hacia la unificación de la electricidad y el magnetismo que logró, unos cincuenta años más tarde, el gran científico escocés James Clerk Maxwell. Hoy día la creencia en la posibilidad de unificar las distintas fuerzas fundamentales de la naturaleza está en su apogeo y a ello dedican sus esfuerzos innumerables científicos, habiéndose obtenido ya resultados muy interesantes.

En la famosa experiencia de Oersted se colocaba un conductor rectilíneo horizontal y debajo de él un imán que podía girar libremente en el plano horizontal. Al hacer pasar por el conductor una corriente eléctrica el imán se movía, situándose perpendicularmente al conductor. Si bien ya se sabía que las cargas eléctricas en reposo no influían sobre los imanes, esta experiencia demostraba que las cargas eléctricas en movimiento de alguna manera actuaban sobre éstos, es decir, las cargas eléctricas en movimiento originaban campos magnéticos. Cuando Oersted anunció su descubrimiento, en julio de 1820, la primera reacción fue de escepticismo, pero pronto éste cambió en entusiasmo y en los meses que siguieron se realizaron varios descubrimientos importantes para comprender la relación entre la electricidad y el magnetismo. En septiembre de ese mismo año, André Marie Ampère explicó la interacción entre dos corrientes y las propiedades magnéticas de los solenoides y desarrolló un modelo teórico que daba cuenta del comportamiento de los imanes permanentes en función de espiras microscópicas de corriente existentes dentro de los materiales magnéticos. Al año justo, Michael Faraday lograba probar que las líneas de fuerza magnética alrededor de un conductor rectilíneo eran círculos concéntricos situados en planos perpendiculares al conductor y centrados en él.

Otro paso fundamental en la búsqueda de la unificación de la electricidad y el magnetismo fue el descubrimiento del fenómeno de inducción electromagnética, es decir, la posibilidad de inducir corrientes eléctricas en circuitos al mover éstos en el seno de un campo magnético o al situarlos en zonas del espacio en las que habían campos magnéticos que variaban en el transcurso del tiempo. Joseph Henry descubría este fenómeno en las vacaciones del verano de 1829, pero no publicaría su resultado hasta tres años más tarde. En este período, y de forma totalmente independiente, Faraday, en 1831, redescubrió el fenómeno, y hoy día se le conoce como el descubridor de la inducción electromagnética no sólo porque estableció la prioridad oficial de su publicación sino también por el estudio exhaustivo que hizo de la misma. Otro paso interesante para nosotros en el desarrollo del electromagnetismo fue el descubrimiento de Faraday, en 1845, que confirmaba la sospechada relación de éste con la luz: demostró que el plano de polarización de la luz giraba al hacerla pasar por un campo magnético.

A mediados del siglo pasado numerosos hechos experimentales confirmaban la íntima relación entre la electricidad y el magnetismo, y se vislumbraba ya una clara relación entre estos fenómenos y la luz. En 1873 se publicó el libro fundamental de Maxwell, *A Treatise on Electricity and Magnetism,* en el que mediante unas pocas ecuaciones fundamentales que relacionaban entre sí el campo eléctrico, el campo magnético y las densidades de carga y de corriente se lograba formular una teoría completa del electromagnetismo, lográndose así una unificación total de la electricidad y el magnetismo *(véase la ilustración de esta página).* Las ecuaciones de Maxwell son para el electromagnetismo el equivalente a las ecuaciones de Newton para la mecánica clásica: a partir de ellas se pueden explicar todos los fenómenos clásicos del electromagnetismo conocidos hoy día. Las ecuaciones de Maxwell predicen también que al acelerar una carga eléctrica ésta emite radiación. Estas ondas electromagnéticas que se propagan con la velocidad de la luz son transversales, es decir, corresponden a perturbaciones perpendiculares a la dirección de propagación. La perturbación que se propaga son variaciones de los campos eléctricos y magnéticos perpendiculares entre sí y a la dirección de propagación del fenómeno. Estas ondas son totalmente idénticas a las de la luz; quedaba así establecida una teoría electromagnética de ésta. La prueba experimental concluyente fue dada en 1887 por Heinrich Rudolph Hertz al generar por vez primera ondas electromagnéticas mediante una bobina de inducción.

Las ecuaciones de Maxwell describen los campos electromagnéticos creados por distribuciones de cargas y corrientes. Nos falta dar una expresión explícita para la fuerza que actúa sobre una partícula de carga q en el seno de un campo electromagnético. Ya hemos dicho antes que si la partícula está en un campo eléctrico **E,** la fuerza que actúa sobre ella es $q\mathbf{E}$. Por otra parte, el campo magnético **B** no actúa sobre cargas en reposo, pero sí lo hace cuando éstas se encuentran en movimiento; si **v** es la velocidad de la carga

$$\nabla \cdot \mathbf{E} = 4\pi\varrho_e \qquad \nabla \times \mathbf{B} = \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c}\mathbf{j}_e$$

$$\nabla \cdot \mathbf{B} = 0 \qquad -\nabla \times \mathbf{E} = \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}$$

ECUACIONES DIFERENCIALES de Maxwell para el campo electromagnético en el vacío. La primera nos permite calcular el campo eléctrico E creado por una distribución de cargas de densidad ϱ_e. La segunda nos da el campo magnético B creado por una densidad de corriente eléctrica $\mathbf{j}_e$ o inducido por variaciones del campo eléctrico en el transcurso del tiempo. La tercera ecuación nos da una propiedad de B que debe satisfacerse por el hecho de no existir cargas magnéticas libres. Finalmente, la última ecuación refleja el hecho de que la variación temporal de un campo magnético crea un campo eléctrico (y, por tanto, una corriente de un conductor) y contiene todas las leyes de la inducción magnética de Faraday. En todas estas ecuaciones, *c* es la velocidad de luz en el vacío y los símbolos ∇ E y ∇ x E (e igual para el campo B) dan unas ciertas combinaciones relacionadas con las variaciones del vector E entre dos puntos próximos del espacio; ∂E/∂t da el ritmo de variación del campo eléctrico E. (Los vectores van en negritas.)

entonces la fuerza es $q\mathbf{v} \times \mathbf{B}/c$, donde c es la velocidad de la luz, y por $\mathbf{v} \times \mathbf{B}$ indicamos un vector que tiene como módulo $\mathbf{v}B$ sen α (siendo α el ángulo que forma la velocidad con el campo magnético), su dirección es perpendicular al plano determinado por **v** y **B** y su sentido es el que corresponde al giro necesario para llevar **v** sobre **B** por el camino más corto. Obtenemos entonces que, para el caso general, la fuerza viene dada por la expresión $q(\mathbf{E} + \mathbf{v} \times \mathbf{B}/c)$, que es la llamada fuerza de Lorentz.

$$\mathbf{E} = -\nabla \Phi - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} \qquad \Phi' = \Phi + \frac{1}{c}\frac{\partial \Lambda}{\partial t}$$

$$\mathbf{B} = \nabla \times \mathbf{A} \qquad \mathbf{A}' = \mathbf{A} - \nabla \Lambda$$

RELACIONES EXPLICITAS entre los potenciales electromagnéticos Φ y A y los campos E y B: se dan en la primera parte de la figura. El último término de la primera ecuación describe la variación del potencial vector en el transcurso del tiempo; para campos estacionarios es posible anular este término, en cuyo caso el potencial escalar no es más que el conocido potencial electrostático. En la segunda parte de esta figura se da una transformación de gauge. Si Φ y A son potenciales correspondientes a un campo electromagnético y Λ es una función arbitraria de la posición y el tiempo entonces Φ' y A' son unos nuevos potenciales que engendran el mismo campo electromagnético. (Los vectores de esta ilustración aparecen en negritas.)

En una descripción clásica del electromagnetismo desempeñan un papel fundamental los campos eléctrico y magnético. Sin embargo, es posible demostrar que se pueden determinar una función escalar $\Phi(\mathbf{r},t)$ y otra vectorial $\mathbf{A}(\mathbf{r},t)$, llamadas respectivamente potencial escalar y potencial vector, tales que, a partir de ellas, se calcule, mediante operaciones matemáticas sencillas, los valores de **E** y **B**. En la ilustración de esta página aparece la forma explícita de estas relaciones. Se puede probar que si se definen los potenciales, y a partir de ellos se calcula el campo electromagnético, los valores de **E** y **B** así obtenidos satisfacen idénticamente dos de las ecuaciones de Maxwell, precisamente las dos últimas de la ilustración de la página anterior; ello no debe sorprender, pues es precisamente la existencia de estas dos ecuaciones lo que permite introducir los potenciales.

Las otras dos ecuaciones de Maxwell dan origen ahora a ecuaciones diferenciales para la función escalar Φ y la función vectorial **A** que relacionan estas cantidades con las densidades de carga y corriente. Aún más, dado un campo electromagnético hay gran libertad en la elección de los potenciales, estando relacionados dos potenciales que den origen a un mismo campo electromagnético mediante las llamadas transformaciones de gauge. Esta gran libertad se puede usar para elegir los potenciales de tal forma que la ecuación que determina el potencial escalar dependa sólo de la densidad de carga y las que determinan el potencial vector sólo de la densidad de corriente; además, la ecuación para Φ y para cada una de las componentes de **A** tienen exactamente la misma estructura matemática. La relación entre Φ y **A** que hay que imponer para que esto ocurra se llama condición de Lorentz. Esta no es suficientemente restrictiva como para fijar los potenciales escalar y vector, los cuales siguen, pues, admitiendo transformaciones de gauge, pero con funciones Λ no arbitrarias.

La ecuación clásica del movimiento de una partícula cargada en el seno de un campo electromagnético se obtiene resolviendo la ecuación fundamental de la mecánica, que establece una proporcionalidad entre la aceleración y la fuerza teniendo en cuenta que la fuerza a que está sometida la partícula es precisamente la fuerza de Lorentz antes considerada y que ésta depende únicamente de los campos eléctrico y magnético. Si bien éstos se pueden describir mediante los potenciales es claro que las trayectorias de la partícula son totalmente independientes del gauge elegido, es decir, de las expresiones que se hayan elegido para los potenciales que describen el campo electromagnético dado. Más aún, lo único realmente importante son los valores de **E** y **B**.

En mecánica cuántica, que rige el comportamiento de todo el microcosmos y de la que la mecánica clásica es una aproximación, la situación es muy distinta. En este caso no se puede hablar de la trayectoria de una partícula, y el máximo conocimiento que podemos tener sobre el comportamiento de una partícula viene determinado por una función compleja $\Psi(\mathbf{r},t)$, llamada función de ondas, cuyo módulo al cuadrado nos da la densidad de probabilidad de encontrar la partícula en el punto de coordenadas **r** en el instante t. En mecánica cuántica, la ecuación de evolución de una partícula cargada es tal que sólo aparecen en ella los potenciales Φ y **A** y no los campos electromagnéticos; sin embargo, es posible mantener la invariancia de la teoría con respecto a las transformaciones de gauge. Para ello es necesario suponer que, bajo una transformación de gauge que pase de $(\Phi, \mathbf{A})$ a $(\Phi', \mathbf{A}')$, la función de ondas $\Psi(\mathbf{r},t)$ pasa a estar multiplicada por una función exponencial resultando $\Psi(\mathbf{r},t)\, exp\,(-iq\Lambda/\hbar c)$ donde i es la unidad imaginaria, Λ es precisamente la función que nos relaciona los potenciales en los dos gauge considerados *(véase la ilustración de esta página)* y $\hbar$ es la constante de Planck dividida por 2π. Si bien vemos que los resultados de la mecánica cuántica son invariantes bajo las transformaciones de gauge una experiencia típicamente cuántica propuesta por Y. Aharonov y D. Bohm, en 1959, y realizada al año siguiente por R. G. Chambers, demuestra claramente que el papel del potencial electromagnético es mucho más importante de lo que podría esperarse en una teoría clásica.

En la ilustración de la página siguiente se muestra el dispositivo experimental necesario para realizar el estudio del efecto Aharonov-Bohm. Aun cuando el campo magnético sea nulo, fuera del solenoide no existe ningún potencial electromagnético que sea nulo en todas las regiones exteriores a éste. Localmente podemos lograr la anulación de los potenciales pero no globalmente, pues la región exterior, suponiendo el solenoide indefinido, no es simplemente conexa. Esto es debido a que existe una relación entre los valores que toma el potencial vector **A** a lo largo de una curva cerrada cualquiera que rodee el solenoide con el flujo del campo magnético a través de cualquier superficie sustentada por esta curva. Como este flujo es no nulo, **A** debe tomar valores no nulos en puntos de cualquier curva cerrada que rodee el solenoide. Se puede probar que lo que dicta el comportamiento de las franjas de interferencia es el factor $exp\,(iq\,\Phi/\hbar c)$, donde Φ es el flujo de campo magnético en el interior del solenoide. Vemos, pues, que, de alguna manera, el electrón es sensible al potencial electromagnético que es lo que no se anula fuera del solenoide, pero de tal forma que se conserva la invariancia gauge de la teoría, pues el valor de Φ es calculable directamente a partir de **B**. Notar que lo realmente importante es todo el factor y no sólo la fase $q\Phi/\hbar c$, pues para dos campos magnéticos distintos tales que los flujos difieren en un múltiplo entero de $2\pi\hbar c/q$, las fases difieren en un múltiplo entero de 2π, pero el factor esencial es idéntico y, así, se observan las mismas figuras de interferencia.

Hay que mencionar que las ecuaciones de Maxwell, juntamente con la ecuación de la fuerza de Lorentz, permiten describir todos los fenómenos electromagnéticos clásicos que conocemos. Más aún, tomando esta teoría como base se puede

construir una teoría de campos cuántica que describe la interacción de los fotones con los electrones, la llamada electrodinámica cuántica, cuyos cálculos teóricos están de acuerdo con los datos experimentales en increíbles órdenes de precisión. Por ejemplo, el momento magnético anómalo del electrón ha sido calculado con una precisión de unas 4 partes en un millón y en este orden de aproximación coincide con los resultados experimentales.

Ya hemos dicho que la teoría del campo electromagnético de Maxwell logra la unificación de la electricidad y el magnetismo, pero es estéticamente muy desagradable el papel totalmente distinto que desempeñan en ella los campos eléctricos y magnéticos. La ausencia de cargas magnéticas libres o monopolos magnéticos, y, por tanto, de corrientes magnéticas originadas por su movimiento, es precisamente la causante de esta asimetría en el comportamiento de **E** y **B**. Desde este punto de vista sería deseable que existieran monopolos magnéticos. Por otra parte J. K. King ha medido la carga eléctrica de las moléculas de hidrógeno, helio y hexafluoruro de azufre en términos de la carga del electrón y ha encontrado resultados compatibles con cero en órdenes de precisión de unas pocas partes en 10^{21}. Teniendo en cuenta la composición de estas moléculas se puede afirmar, con un orden de precisión análogo, que el neutrón tiene carga nula y que las cargas del protón y del electrón son iguales y opuestas. Si bien con precisión menor se sabe que todas las partículas conocidas tienen cargas eléctricas que son múltiplos enteros de la carga del electrón, es decir, que la carga eléctrica está cuantificada.

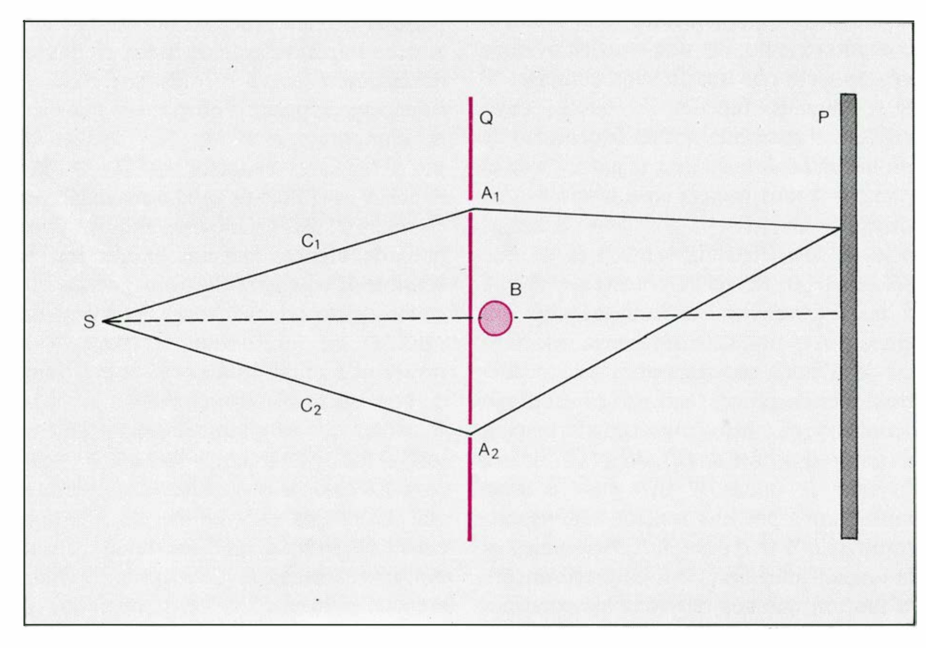
Hoy día se cree que los hadrones están formados por quarks cuyas cargas, en el modelo de Gell-Mann-Zweig, son $2/3$ y $-1/3$ de la carga del protón. La existencia de estas partículas, que aún no han sido encontradas en forma libre y que se sospecha que pudieran existir sólo confinadas en el interior de los hadrones, no invalida la afirmación de que la carga eléctrica esté cuantificada, sino que únicamente deberíamos tomar como carga eléctrica elemental un tercio de la carga del electrón.

En 1931, P. A. M. Dirac se dio cuenta de que la existencia de monopolos no sólo haría que las ecuaciones de Maxwell fueran más simétricas sino que tal existencia, aunque fuera de un solo monopolo en el Universo, permitiría dar una explicación de la cuantificación de la carga eléctrica. La existencia de los monopolos permitiría, pues, dar dos pasos importantes hacia adelante en la comprensión de los fenómenos electromagnéticos.

Supongamos que los monopolos existan; entonces, podremos introducir, de forma totalmente análoga a como introdujimos la densidad de carga y de corriente eléctrica, una densidad de carga y de corriente magnética que designaremos por ρ_m y $\mathbf{j}_m$, respectivamente. Las ecuaciones de Maxwell modificadas vienen dadas en la ilustración de la página siguiente. De la misma forma que las ecuaciones de Maxwell incorporaban la ecuación de continuidad que garantiza la conservación de la carga eléctrica, estas ecuaciones implican que ρ_m y $\mathbf{j}_m$ satisfacen una ecuación análoga que implica la conservación de la carga magnética. Por otra parte, debemos aún especificar la ley de fuerza que actúa sobre una partícula puntual con carga eléctrica q y carga magnética g en el seno de un campo electromagnético. La generalización natural de la ley de Lorentz dada antes hace suponer que dicha fuerza tiene la expresión $q\,(\mathbf{E} + \mathbf{v}\times\mathbf{B}/c) + g\,(\mathbf{B} - \mathbf{v}\times\mathbf{E}/c)$. En relación con esta ley, vale la pena mencionar que una partícula con sólo carga eléctrica no gana energía al moverse en el seno de un campo magnético, mientras que un monopolo que recorra una distancia L, moviéndose paralelamente y en el mismo sentido de un campo magnético constante y uniforme, gana una energía gBL. N. A. Porter ha sugerido que los rayos cósmicos de energías muy altas (mayores que 10^{18} eV) quizá sean monopolos que han sido acelerados por el campo magnético de nuestra galaxia. Mencionemos sólo que ésta no es la única explicación; más adelante volveremos sobre esta hipótesis.

Por otra parte, si no existen monopolos el electromagnetismo es invariante bajo las operaciones de paridad y de inversión temporal. Una consecuencia de la invariancia de la teoría bajo la transformación de paridad es que si $\mathbf{r}(t)$ es una trayectoria posible de una partícula en el seno de un campo electromagnético $\mathbf{E}(\mathbf{r},t)$, $\mathbf{B}(\mathbf{r},t)$, entonces es también una trayectoria posible $-\mathbf{r}(t)$ en el seno del campo $-\mathbf{E}(-\mathbf{r},t)$, $+\mathbf{B}(-\mathbf{r},t)$. La invariancia bajo inversión temporal implica que también es trayectoria posible $\mathbf{r}(-t)$ en el seno del campo $+\mathbf{E}(\mathbf{r},-t)$, $-\mathbf{B}(\mathbf{r},-t)$. De existir los monopolos magnéticos, esto no es cierto y, por tanto, su existencia implica que las interacciones electromagnéticas violan ambas invariancias. Sabemos que estas invariancias no son válidas en el marco de las partículas elementales, pero los resultados actuales indican que estas violaciones están asociadas, de alguna forma, con las interacciones débiles, las causantes de las desintegraciones beta de los núcleos atómicos. Quizás en el futuro se las pueda relacionar con las interacciones electromagnéticas y utilizar los monopolos como el vehículo para violar dichas simetrías.

Hemos dicho que en las ecuaciones de Maxwell modificadas los campos eléctrico y magnético desempeñan un papel simétrico. Esto se traduce en el hecho de que dichas ecuaciones son invariantes bajo lo que se suele denominar una transformación de dualidad. Indiquemos por las cantidades $(\mathbf{a}_e, \mathbf{a}_m)$ cada una de las parejas $(\mathbf{E}, \mathbf{B})$, (ϱ_e, ϱ_m) y $(\mathbf{j}_e, \mathbf{j}_m)$, entonces las transformaciones de dualidad serán el



LA FUENTE S **produce electrones que después de atravesar la pantalla** Q **por las rendijas** A_1 **y** A_2**, perpendiculares al plano de la figura, originan en la pantalla** P **las típicas figuras de interferencia debidas a la onda asociada al electrón. La novedad consiste en un solenoide colocado en** B **tal que el campo magnético, confinado totalmente en su interior, sea perpendicular al plano del dibujo, siendo el tamaño de este solenoide lo suficientemente pequeño para que los electrones tengan probabilidad despreciable de penetrar en su interior. Al aumentar el campo magnético en el interior del solenoide se ve cómo la figura de interferencia producida en la pantalla va cambiando, aun cuando los electrones, al ser la región** B **suficientemente pequeña, nunca pasan por la región del espacio en la que el campo magnético es distinto de cero.**

conjunto de transformaciones que nos hacen pasar de (a_e, a_m) a $(a'_e = a_e \cos\delta + a_m \operatorname{sen}\delta,\ a'_m = -a_e \operatorname{sen}\delta + a_m \cos\delta)$, donde δ es una cantidad real arbitraria. Es fácil darse cuenta que estas transformaciones entre el campo eléctrico y el magnético, entre las densidades de carga eléctrica y magnética y entre las densidades de corriente eléctrica y magnética son tales que las ecuaciones de Maxwell generalizadas antes y después de la transformación tienen exactamente la misma forma, es decir, son invariantes. No sólo esto, sino que bajo estas transformaciones también permanecen invariantes las expresiones de la fuerza de Lorentz y todas las cantidades características de un campo electromagnético, tales como las densidades de energía, impulso y momento angular.

Una consecuencia importante e inmediata de la invariancia de la teoría bajo las transformaciones de dualidad es la siguiente: Supongamos que un determinado observador constatara que todo el universo está formado por un único tipo de partículas que tienen una carga eléctrica q y una carga magnética g, que crean y son movidas por unos campos electromagnéticos **E** y **B**. Evidentemente observará que la densidad de carga eléctrica partida por la densidad de carga magnética es igual a q/g, y una relación análoga para las corrientes. Entonces, haciendo una transformación de dualidad tal que $\operatorname{sen}\delta = g/\sqrt{q^2 + g^2}$ y $\cos\delta = q/\sqrt{q^2 + g^2}$ se obtendría una descripción totalmente equivalente en la que las partículas tendrían carga eléctrica $q' = \sqrt{q^2 + g^2}$ y carga magnética $g' = 0$. Esto es, si sólo hubiera un tipo de partículas en el universo, se podría fijar de forma totalmente arbitraria la relación entre sus cargas eléctrica y magnética y sería siempre posible hablar únicamente de cargas eléctricas. En el mundo real hay distintas clases de partículas; el razonamiento anterior nos permite ver que se puede fijar arbitrariamente como cero la carga magnética de un determinado tipo de partículas, por ejemplo, los electrones. Entonces tiene pleno sentido preguntarnos cuáles son las cargas magnéticas de todas las restantes partículas tales como protones, neutrones,... Más adelante discutiremos eseste punto.

En su teoría, Dirac suponía que una partícula podía tener carga eléctrica o carga magnética, pero no ambos tipos de carga simultáneamente. Consideremos una partícula de masa m y carga q que se mueve lentamente en el campo magnético creado por un monopolo de carga magnética g, que está fijo en un punto del espacio, que tomaremos como origen de coordenadas. De la misma forma que una partícula de carga q, situada en el origen de coordenadas, crea un campo eléctrico $\mathbf{E}(\mathbf{r}) = q\mathbf{r}/r^3$, este monopolo creará un campo magnético $\mathbf{B}(\mathbf{r}) = g\mathbf{r}/r^3$ y, debido a él, la partícula de carga q siente la acción de una fuerza que, de acuerdo con la expresión de Lorentz, es $q\mathbf{v} \times \mathbf{B}/c$. El campo magnético creado por el monopolo tiene simetría esférica y, de acuerdo con teoremas generales que asocian a simetrías continuas leyes de conservación, es de esperar que exista una cantidad vectorial que se denomina momento angular total **J**, que se conserve en el transcurso del tiempo, es decir, que ni el módulo ni la dirección de **J** cambian con el tiempo. Esta cantidad conservada resulta ser la suma de dos términos, uno de los cuales corresponde al momento angular orbital de la partícula con carga eléctrica y que es $m\mathbf{r} \times \mathbf{v}$, donde **v** simboliza la velocidad de esta partícula; el otro es el momento angular asociado con el campo electromagnético presente en el sistema cuyo valor es $-qg\mathbf{r}/rc$. En particular podemos calcular la proyección del momento angular total en la dirección de la línea que une el monopolo con la carga eléctrica y ésta resulta ser $-qg/c$. Al establecer la estructura matemática de la mecánica cuántica a todo observable se le hace corresponder un objeto que se denomina operador autoadjunto; dados dos operadores de este tipo, A y B, su conmutador, definido como $AB\text{-}BA$ es, en general, distinto de cero. Si se acepta que los conmutadores de los operadores asociados con los observables posición y velocidad de una partícula cargada en el seno de un campo magnético son aun válidos cuando éste es debido a un monopolo, se puede ver que el momento angular total dado antes tiene un operador asociado que cumple las reglas de conmutación típicas del momento angular en mecánica cuántica. Esto implica que el valor de la proyección del momento angular en una dirección cualquiera del espacio sólo puede tomar valores que son múltiplos enteros de $\hbar/2$ ($\hbar = h/2\pi$, siendo h la constante de Planck). Teniendo en cuenta que, como ya hemos explicado, la proyección del momento angular total en la dirección de la recta que va desde el monopolo a la partícula vale $-qg/c$, el resultado anterior conduce inmediatamente a la ley de cuantificación de Dirac, a saber: $qg = (n/2)\hbar c$.

Argumentos simples de teoría de números permiten obtener a partir de la condición de Dirac el siguiente resultado: todas las cargas eléctricas deben ser múltiplos enteros de una carga eléctrica elemental q_{el} y todas las cargas magnéticas deben ser múltiplos enteros de una carga magnética elemental g_{el}. Por otra parte estas cargas elementales, que por razonamientos análogos se demuestra que son físicamente realizables, no resultan arbitrarias, sino que deben satisfacer la relación $q_{el} \cdot g_{el} = n\hbar c/2$, donde n es un número entero arbitrario no especificado por la teoría. Si se toma como carga eléctrica elemental la del electrón (y así lo haremos de no decir específicamente lo contrario) y se introduce la llamada constante de estructura fina α_e, que se define como $e^2/\hbar c$, es una constante adimensional que mide la intensidad de las interacciones electromagnéticas y que vale aproximadamente $1/137$), entonces la cantidad magnética correspondiente es $\alpha_m = n^2/4\alpha_e \simeq 34\ n^2$ que es muchísimo mayor que α_e, incluso en el caso en que $n = 1$. Evidentemente, si se tomara como carga elemental $1/3$ de la del electrón, la unidad de carga magnética sería aún mayor. Vemos pues que al introducir los monopolos logramos una simetría teórica entre la electricidad y el magnetismo; sin embargo, teniendo en cuenta el valor experimental de la constante de estructura fina se observa que las fuerzas magnéticas son mucho más intensas que las eléctricas, lo cual hace surgir una asimetría desde el punto de vista práctico.

J. Schwinger ha considerado también la

$$\nabla \cdot \mathbf{E} = 4\pi \varrho_e \qquad \nabla \times \mathbf{B} = \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c}\mathbf{j}_e$$

$$\nabla \cdot \mathbf{B} = 4\pi \varrho_m \qquad -\nabla \times \mathbf{E} = \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} + \frac{4\pi}{c}\mathbf{j}_m$$

a *b*

EXTENSION NATURAL de las ecuaciones de Maxwell cuando se supone que existen monopolos magnéticos. Estas ecuaciones serían totalmente simétricas bajo el intercambio de las magnitudes eléctricas E, ϱ_e y j_e con las magnéticas B, ϱ_m y j_m si no fuese por el signo menos que aparece en la última de ellas y que hace que la transformación de simetría que las deja invariantes sea sustituir las magnitudes eléctricas por las magnéticas y las magnéticas por las eléctricas cambiadas de signo. Esta simetría, que de hecho es un caso particular de una más general discutida en el texto, ya indica que en esta teoría el que un campo, carga o corriente sea eléctrico o magnético dependerá, en última instancia, de una definición de un cierto campo, carga o corriente como eléctrico o magnético. Ahora ya no es posible definir unos potenciales escalar y vector según las expresiones dadas en la primera figura de la ilustración de la página 7, pues éstas satisfacían idénticamente las dos últimas ecuaciones de Maxwell, que son las modificadas al introducir los monopolos.

posibilidad de que existan dyones, es decir, partículas que tengan simultáneamente cargas eléctricas y magnéticas. Siguiendo razonamientos análogos a los de Dirac se obtiene que dados dos dyones con cargas (q_1,g_1) y (q_2,g_2), éstas deben ser tales que $q_1g_2-q_2g_1$ sea un múltiplo entero de $\hbar c/2$. Un análisis simple de esta condición muestra que si la carga magnética elemental es $g_{el} = n\hbar c/2|e|$, donde e es la carga del electrón, entonces todas las partículas compuestas que no tengan carga magnética deben poseer una carga eléctrica que sea un múltiplo entero de la del electrón, mientras que la carga eléctrica de aquellas que también tienen carga magnética está indeterminada. Sin embargo, si se supone además que existen partículas compuestas eléctricamente neutras pero con carga magnética, entonces las partículas con carga magnética tienen cargas eléctricas que son fracciones racionales de la del electrón. Ya que en el modelo de Gell-Mann-Zweig los quarks tienen cargas fraccionarias se ha sugerido que los quarks son dyones y se han construido modelos de hadrones en los que este hecho ha sido tenido en cuenta.

Si consideramos el campo magnético creado por un monopolo, no se satisface la tercera ecuación de Maxwell y entonces es imposible definir en todo el espacio el potencial vector **A**; se prueba que, a lo sumo, se puede definir un potencial vector que esté bien definido en todo el espacio salvo en una línea que una la posición del monopolo con el infinito. Una forma de lograr esto es considerar un solenoide rectilíneo de longitud infinita cuyo polo positivo, que tiene intensidad g, está colocado en el origen de coordenadas. Hagamos que la sección del solenoide tienda hacia cero, en cuyo caso el campo magnético en el interior del solenoide tiende hacia infinito, pues el flujo de este campo a través de una sección recta debe valer $4\pi g$. El campo magnético del solenoide $\mathbf{B}_{sol}$ se puede describir mediante un potencial vector **A** de tal forma que en todos los puntos del espacio se cumple que $\mathbf{B}_{sol} = \nabla \times \mathbf{A}$; por otra parte, este campo magnético coincide con el creado por un monopolo de intensidad g colocado en el origen, salvo en la línea ocupada por el solenoide. Entonces podemos describir el campo del monopolo mediante el campo magnético creado por el solenoide, y que puede ser descrito por un potencial vector **A**, restándole el campo magnético en el interior del solenoide *(véase la ilustración de la página siguiente)*. La línea ocupada en el solenoide se llama "cuerda de Dirac".

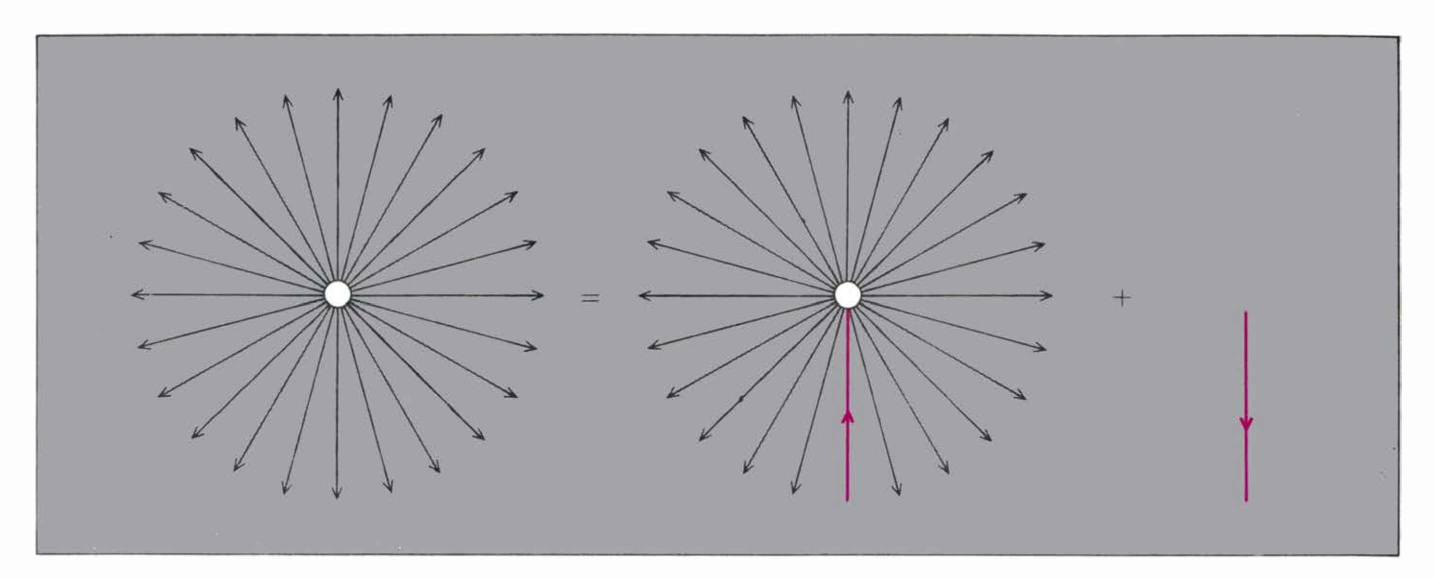
Fijada la posición de la cuerda de Dirac, resulta fácil calcular un potencial vector que nos dé el campo magnético creado por el monopolo en todos los puntos del espacio salvo en dicha cuerda, basta tomar un potencial que nos dé $\mathbf{B}_{sol}$, que presenta una singularidad sobre la cuerda originada por el valor infinito que tiene en ella el campo magnético. Dirac sugirió que se usara precisamente este potencial vector para establecer la ecuación mecánico-cuántica del movimiento de una partícula cargada en el campo creado por el monopolo. Con esto y el formalismo cuántico se obtiene, al exigir que la función de onda que describe la partícula cargada sea univaluada (es decir, que tenga un valor bien definido en cada punto del espacio), la condición de cuantificación de Dirac de una forma más satisfactoria. Se debe hacer notar que el flujo a lo largo de la cuerda de Dirac, al imponer la condición de cuantificación, es tal que $q\Phi/\hbar c$ es un múltiplo entero de 2π, que es la misma condición que se obtenía para que la presencia de un solenoide no modificara la figura de interferencia en la experiencia de Aharonov-Bohm. El campo magnético de un monopolo difiere, pues, del campo del solenoide rectilíneo e indefinido que se ha usado para definir el potencial vector en un tubo de flujo no observable si se cumple la condición de cuantificación de Dirac; por tanto, el cambio del campo magnético del monopolo por el campo magnético del solenoide corresponde a un cambio no observable en tanto que la partícula no cruce la cuerda de Dirac.

Para que toda esta descripción sea coherente es necesario probar que las ecuaciones mecánico-cuánticas que resultan de ella son independientes del cambio de la orientación del solenoide rectilíneo o de la deformación de éste, en una línea cualquiera que una el origen con el punto del infinito. Evidentemente, si hacemos una transformación de gauge en el que la función Λ es regular y univaluada en todo el espacio, el potencial creado por el solenoide no cambia y, por tanto, tampoco el término que describe la cuerda de Dirac. Para pasar de una posible posición de la cuerda S a otra S', que difiere de S en una línea de longitud finita, debemos ampliar el concepto de transformación de gauge.

Sea C la curva que partiendo del origen sigue la línea S hasta llegar a un punto de intersección con S' y que vuelve al origen siguiendo S'. Podemos ahora imaginar una superficie que venga sustentada por C; desde un punto del espacio de coordenadas **r** no perteneciente a la superficie veremos ésta bajo un ángulo sólido Ω (**r**), claramente imposible de definir si el punto pertenece a la curva C. Este ángulo sólido se define de tal forma que diferentes elecciones de la superficie conducen a valores de Ω (**r**) que difieren en múltiplos enteros de 4π *(véase la ilustración de la página 12)*. Se puede probar que la transformación de gauge multivaluada $\Lambda = g\Omega$ (**r**) relaciona las dos distintas elecciones de la cuerda de Dirac consideradas. Bajo esta transformación de gauge la función de onda adquiere, según ya hemos dicho, un factor de fase $exp\ (-iqg\Omega/\hbar c)$ y, por tanto, el hecho de que la función de onda sea univaluada implica que este factor resulte independiente de la superficie elegida; es decir, no cambia al variar Ω en un múltiplo entero de 4π. Ello nos conduce de nuevo, con un grado mayor de rigor y abstracción, a la tantas veces mencionada condición de cuantificación de Dirac.

Si en un punto del espacio consideramos dos transformaciones de gauge sucesivas, engendradas por las funciones Λ_1 y Λ_2, que al fijar el punto del espacio se convierten en números reales, la función de ondas cambia en un factor $exp\ [-iq\ (\Lambda_1 + \Lambda_2)/\ \hbar c]$, cambio que es idéntico al que se hubiera obtenido al realizar una transformación de gauge con $\Lambda = \Lambda_1 + \Lambda_2$, lo cual permite probar que las transformaciones de gauge forman un grupo. De no existir cuantificación de la carga eléctrica al considerar partículas distintas encontraríamos para sus funciones de ondas factores de transformación distintos para cada valor de Λ, lo cual nos dice que el grupo de transformaciones de gauge es equivalente al determinado por los números reales bajo la operación de suma y que los matemáticos denominan grupo no compacto T ($\mathbb{R}$). Si, por el contrario, la carga eléctrica está cuantificada, entonces los únicos gauges físicamente distinguibles, es decir, que conducen a factores de fase distintos, son aquéllos contenidos en el intervalo $[0,\ 2\pi\hbar c/q_{el}]$, pues toda otra transformación de gauge es totalmente equivalente a otra que yace en este intervalo. Debido a esta equivalencia, el grupo de transformaciones de gauge es ahora totalmente análogo al grupo de rotaciones alrededor de un eje que, en el argot matemático, se denomina grupo compacto U(1).

Hemos visto en la última parte la creciente importancia de las transformaciones de gauge en orden a establecer, con rigor creciente, la cuantificación de la carga eléctrica y hemos terminado viendo la relación íntima entre este fenómeno y la topología global del grupo de gauge. Por otra parte, en los esquemas de unificación de las interacciones electromagnéticas y débiles que se han formulado en estos últimos años son precisamente grupos de transformaciones de gauge, aunque de estructura matemática más complicada que la asociada al campo electromagnético, las que han desempeñado un papel fundamental. También parece vislumbrarse, cada día con mayor claridad, que son precisamente estos grupos los que pueden ayudar a una total comprensión de las interacciones fuertes y



EN LA PARTE DE LA IZQUIERDA de la figura se representa, mediante líneas de fuerza, el campo magnético creado por un monopolo de carga magnética *g*. Este campo magnético es igual al creado por un solenoide rectilíneo indefinido cuyo polo positivo tenga intensidad *g* y se encuentre en la posición del monopolo, en todos los puntos del espacio salvo en el interior del solenoide, indicado por la línea roja en el dibujo. En el interior del solenoide, cuya sección recta β es nula, el campo magnético es infinito de forma que el flujo por su interior es $4\pi g$. Para poder igualar ambos campos es necesario añadir al campo del solenoide un campo magnético igual y opuesto al que existe en su interior. La línea ocupada por el solenoide se denomina "cuerda de Dirac".

gravitacionales y a una futura unificación de todas las interacciones fundamentales. Un ejemplo de un grupo ampliamente usado en este contexto es el grupo compacto SU(2), el grupo formado por todas las matrices dos por dos que son unitarias y cuyo determinante vale la unidad y cuya ley de multiplicación es la multiplicación ordinaria de matrices. Este grupo coincide localmente, es decir, en un entorno del elemento unidad, con el grupo que describe las rotaciones alrededor de un punto en un espacio tridimensional. Si el operador mecánico-cuántico asociado a la carga eléctrica, cuyos valores propios son las cargas posibles de las partículas que intervienen en la teoría, es proporcional a uno de los generadores del grupo, siendo estos los operadores equivalentes al momento angular en el grupo de rotaciones, entonces de la misma forma que el momento angular está cuantificado quedan cuantificadas las cargas eléctricas.

A primera vista, parece que así podemos obtener un procedimiento alternativo para cuantificar la carga eléctrica sin necesidad de introducir monopolos magnéticos. Esto es engañoso, pues es precisamente en este tipo de modelos en los que en 1974 G. 't Hooft y A. M. Polyakov, de forma independiente, demostraron la existencia de soluciones análogas a los monopolos. La ventaja de esta forma de proceder es que los monopolos no deben ser introducidos *ad hoc*, sino que aparecen como consecuencias lógicas de la teoría. El monopolo de 't Hooft-Polyakov es estable, tiene una carga magnética correspondiente a un monopolo de Dirac con $n = 1$. En principio, se puede calcular su tamaño, ya que no son puntuales, y su masa en términos de las partículas de gran masa que suelen aparecer en estas teorías; lo cual implica que estos monopolos tienen probablemente masas bastante superiores a 1000 veces la masa del protón y, por tanto, de ser esto así, no es posible producirlos con los aceleradores actuales. Así pues, y hasta hoy día, no se ha dado con una razón para la cuantización de la carga eléctrica que de forma directa o indirecta no implique la existencia de monopolos.

Antes de terminar esta breve exposición de resultados teóricos queremos comentar que la existencia de monopolos tiene otra consecuencia de gran importancia. De acuerdo con la mecánica cuántica, el spin (momento angular intrínseco) de una partícula elemental puede ser entero o semientero (usaremos como es habitual e incorrecto esta palabra para indicar un número semi-impar 1/2, 3/2, 5/2,...) Para agregados de partículas se suele denominar spin al momento angular total del mismo referido a su centro de masa. Existe un importante teorema, llamado de conexión spin-estadística, que afirma que todas las partículas o agregados de éstas de spin entero son bosones y que las de spin semientero son fermiones. Por bosones y fermiones entendemos que los sistemas obedecen, respectivamente, a las estadísticas de Bose-Einstein y de Fermi-Dirac.

Este hecho implica comportamientos muy distintos; así, por ejemplo, para los fermiones existe un principio de exclusión que nos permite afirmar que nunca se pueden encontrar dos partículas idénticas en el mismo estado cuántico, mientras que esta prohibición no existe para los bosones. Por otra parte, de las leyes de suma de momentos angulares en mecánica cuántica se deduce que, si bien un agregado con un número par de fermiones es un bosón, no hay forma de construir un fermión como agregado de bosones. El análisis de Dirac muestra que un agregado formado por un monopolo y una carga, ambos sin spin, tiene spin semientero debido a la contribución del campo electromagnético creado por el sistema monopolo-carga.

Un análisis superficial del problema implicaría que estos agregados, a pesar de su spin, fueran bosones, violándose el teorema de conexión spin-estadística. La realidad es muy otra; A. S. Goldhaber, tras un examen cuidadoso del problema, ha demostrado que es precisamente la interacción de largo alcance entre cargas y monopolos de distintos agregados lo que restablecía la correcta conexión entre el spin y la estadística. Si existen monopolos es posible construir fermiones a partir de bosones. Resulta extraordinario observar cómo la simetría entre la electricidad y el magnetismo implica también una simetría entre las dos estadísticas cuánticas, haciendo que fermiones puedan estar formados por sólo bosones, de forma análoga a como bosones pueden construirse con sólo fermiones.

Después de haber pasado revista a la situación teórica queremos analizar, aunque sea brevemente, la experimental. La primera pregunta que nos podemos formular es si las partículas elementales, descubiertas, tienen carga magnética. H. Harrison y colaboradores han encontrado una forma sencilla para establecer una cota superior a la carga magnética de los nucleones (neutrones y protones): La Tierra está formada por nucleones y electrones, que son los componentes básicos de todos los núcleos atómicos, residiendo prácticamente toda su masa en los nucleones. Sabiendo que el radio de la Tierra es $R \simeq 6{,}4 \times 10^8$ centímetros, aproximadamente, y conociendo su masa es fácil de-

terminar que contiene unos $N \simeq 3{,}4 \times 10^{51}$ nucleones. Supongamos que cada uno de ellos tuviera una carga magnética g_n y elijamos, cosa siempre posible, que la carga magnética de los electrones sea nula, entonces los nucleones crearían en la superficie terrestre un campo magnético $|g_n| N/R^2$; como sabemos que este campo es menor de 1 gauss se deduce inmediatamente que $|g_n| < 2 \times 10^{-25}\,|e|$, donde $|e|$ es la carga eléctrica del protón.

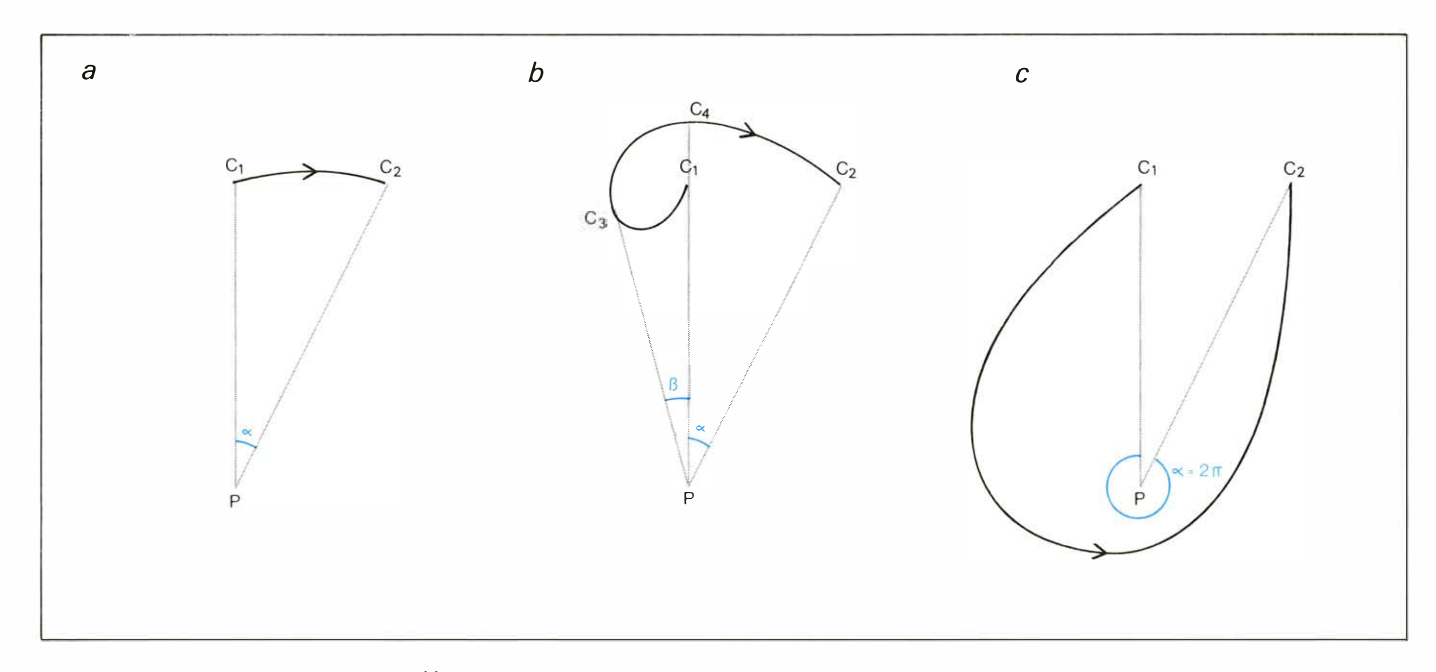
Se podrían poner algunas objeciones a este razonamiento (las cargas magnéticas del protón compensarán en todo o en parte la de los neutrones, las corrientes internas terrestres compensarán en parte el campo creado por los monopolos, etcétera.), pero un análisis parecido de otros cuerpos celestes permite llegar a conclusiones similares. Más aún, L. L. Vant-Hull, aprovechando el hecho de que si las partículas tuvieran cargas monopolares magnéticas al moverse y cruzar el plano de una espira inducirían en ella una fuerza electromotriz inducida de forma totalmente análoga a como al variar el flujo magnético que atraviesa una espira se induce en ella una fuerza electromotriz, ha encontrado que la carga magnética del protón y del neutrón deben ser menores que $1 \times 10^{-26}\,|e|$. Estos resultados y otros parecidos que, con menos precisión, pueden obtenerse para las partículas elementales más usuales conducen a la creencia de que todas las partículas elementales conocidas carecen de carga magnética.

En los laboratorios y en la industria se usan frecuentemente campos magnéticos que se crean a partir de corrientes eléctricas; las propiedades magnéticas de la materia pueden explicarse como debidos a las corrientes eléctricas que crean los electrones en su movimiento. Muchas partículas elementales tienen momentos dipolares magnéticos, es decir, actúan como pequeños imanes, y podemos preguntarnos si son debidos a corrientes eléctricas internas de estas partículas o si por el contrario existen en su interior pares de monopolos con cargas magnéticas opuestas que, aunque determinan que su carga magnética total sea nula, crean un momento dipolar magnético.

Esta cuestión fue elegantemente analizada por J. D. Jackson, en 1977, y ha probado, basándose en el estudio de la diferencia de energía entre ciertos niveles atómicos llamada estructura hiperfina y de los procesos de colisión de neutrones en hierro, que para el protón, el neutrón, el electrón y el muón la parte de sus momentos dipolares magnéticos creados por monopolos es a lo sumo una cienmilésima del total, siendo el resto creado por corrientes eléctricas. Más aún, como la introducción de monopolos implica que paridad e inversión temporal dejan de ser transformaciones de simetría del electromagnetismo, las partículas elementales pueden tener momentos dipolares eléctricos, que estaban precisamente prohibidos sin estas simetrías.

Ahora bien, experimentalmente se sabe que de existir éstos, deben ser extraordinariamente pequeños, lo cual conduce a la necesidad de que si hay algunas cargas magnéticas en su interior deben estar muy difuminadas de forma que la densidad de carga magnética efectiva sea extraordinariamente pequeña en cada punto de la partícula, lo cual crea problemas al intentar construir modelos de hadrones en los que los quarks sean dyones. Las cargas magnéticas pueden existir libres o formando pares ligados, pero, si existen, no tienen nada, o casi nada, que ver con las propiedades magnéticas de los medios materiales o de las partículas elementales. Los fenómenos magnéticos que se observan son debidos a corrientes eléctricas.

Desde hace años se viene realizando cierto esfuerzo en la búsqueda de monopolos que puedan existir en la naturaleza o en crear éstos mediante aceleradores de partículas y proceder después a su detección. En un estudio de todas estas experiencias, que no pretendemos hacer aquí, es necesario tener en cuenta dos criticismos: (1) Las propiedades de los monopolos a grandes distancias, debidas a sus interacciones electromagnéticas, pueden obtenerse, en un nivel de confianza satisfactorio, aplicando las transformaciones de dualidad a las correspondientes propiedades de las cargas eléctricas y teniendo en cuenta las diferencias en cargas y masas. (2) El comportamiento de los monopolos en el interior de la materia cuesta más predecirlo en razón de posibles peculiaridades de su interacción de corto alcance con los núcleos atómicos; esto hace particularmente difícil calibrar la importancia de experiencias con resultados negativos en los que los monopolos se han difundido lentamente a través de la materia o en las que se han intentado extraer de materiales magnéticos mediante campos magnéticos intensos.



DEFINICION DEL ANGULO SOLIDO $\Omega(r)$ en el caso de un espacio bidimensional. Los dos puntos C_1 y C_2 corresponden a la curva cerrada C. La línea que une C_1 y C_2 corresponde a la superficie sustentada por C. El ángulo bajo el que se ve esta línea desde P corresponde a $\Omega(r)$. Este ángulo se mide recorriendo la línea desde C_1 a C_2, siendo positivos en sentido horario y negativo en sentido contrario. En *a)* vale α. En *b)* también vale α, pues las contribuciones al ir de C_1 a C_3 y de C_3 a C_4 son $-\beta$ y β, respectivamente, cancelándose entre sí. En *c)* vale $\alpha - 2\pi$, donde el 2π corresponde al 4π en el espacio tridimensional.

Se han realizado experiencias en las que se pretenden crear pares monopolo-antimonopolo y posteriormente detectarlos mediante los grandes aceleradores de protones del Fermilab (Estados Unidos) y de Sherpukov (Unión Soviética), mediante los ISR de la organización europea CERN (Suiza), en los que chocan dos haces de protones y el SPEAR de SLAC (Estados Unidos) en el que se hacen chocar un haz de electrones con otro haz de antielectrones. Mencionaremos, como ejemplo, la experiencia llevada a cabo por R. A. Carrigan y colaboradores, en el Fermilab: no han encontrado ningún monopolo con cargas entre 1/30 y 24 veces la carga de un monopolo de Dirac con $n = 1$; lo cual ha permitido establecer que la sección eficaz de producción de monopolos en las colisiones protón-nucleón debe ser menor que 5×10^{-42} cm^2, valor extraordinariamente pequeño. Resultados negativos análogos han sido obtenidos en muchas otras experiencias. A varias de estas experiencias se les puede aplicar el criticismo (2) presentado antes. Por otra parte, si en estos grandes aceleradores el haz de protones tiene una energía de E GeV, sólo es posible crear con ellos monopolos cuando su masa sea menor que $\sqrt{E/2}$ GeV. Si su masa sobrepasara sólo en unas veinte veces la masa del protón sería imposible crear monopolos con los aceleradores de que disponemos ahora. Esta masa es mucho menor que la masa que se espera para los monopolos que aparecen en las teorías gauge.

Por otra parte se han realizado extensas búsquedas de monopolos en la naturaleza. La idea fundamental es que rayos cósmicos muy energéticos son capaces de producir pares monopolo-antimonopolo en la alta atmósfera o que incluso algunos de estos rayos pueden ser monopolos de acuerdo con la hipótesis de Porter; estos monopolos se difunden y quedan atrapados en rocas o dejan trazas, si $n > 1$, en materiales tales como micas u obsidianas. Hasta ahora todos los resultados son negativos, y ello permite establecer límites sobre la concentración de monopolos en distintos medios materiales; tales límites son de menos de un monopolo en 10^{28}-10^{30} nucleones. En muchas de estas experiencias se intentan arrancar mediante campos magnéticos monopolos atrapados en la materia; por consiguiente, se hallan sujetos al criticismo (2). Mencionaremos como particularmente interesantes las experiencias llevadas a cabo por P. H. Eberhard y colaboradores que han buscado monopolos en materiales procedentes de la superficie lunar, sometida durante muchos años a la radiación cósmica, aprovechando la fuerza electromotriz inducida que genera en una espira una muestra en movimento que contenga monopolos. Han encontrado que la máxima densidad posible de monopolos parece ser muy baja: del orden del 8×10^{-28} monopolos por nucleón. De sus resultados pueden deducirse límites sobre el flujo de monopolos en la radiación cósmica y sobre la sección eficaz de creación de pares de monopolos en las colisiones protón-nucleón.

Ya hemos dicho repetidamente que no disponemos todavía de ninguna prueba directa de la existencia de monopolos magnéticos, y, por tanto, sería interesante intensificar su búsqueda. Si los resultados continúan siendo negativos podemos tomar dos posiciones: aceptar que los monopolos magnéticos no existen o aceptar que existen, pero dar una explicación convincente del fracaso de las experiencias realizadas.

Siguiendo la segunda opción una explicación simple sería suponer que las cargas magnéticas son dyones, identificables con los quarks; el posible confinamiento de éstos implicaría el confinamiento de las cargas magnéticas. Otra posible explicación ha sido sugerida por M. A. Ruderman y D. Zwanziger: basándose en la validez para distancias mayores que la longitud de onda Compton asociada al hadrón más ligero (el pión), de una ley de atracción entre un monopolo y un antimonopolo totalmente idéntica a la ley de Coulomb entre cargas eléctricas, obtienen una cota inferior para las masas de los monopolos que es varias veces la masa del protón. Esto implica que, a no ser que su masa sea muy grande, en cuyo caso su producción requeriría una gran energía, el par monopolo-antimonopolo tenga que ser altamente relativista para poder escapar a su propia fuerza atractiva. Pero entonces el hecho de que su carga magnética sea grande hace que estos monopolos radien muy intensamente y que antes de poder separarse hayan perdido prácticamente su energía por radiación aniquilándose finalmente. Estiman que su vida media sería de unos 10^{-21}–10^{-22} segundos y que el único residuo visible que quedaría de todo el proceso sería un haz con muchos fotones, que constituirían la prueba experimental indirecta de su existencia. Si bien en placas nucleares sometidas a radiación cósmica se han visto algunos sucesos que podrían ser interpretados como acabamos de explicar, no se consideran tampoco prueba fehaciente de la existencia de monopolos.

Los monopolos simetrizan totalmente el electromagnetismo, cuantifican la carga, pueden ser portadores de cargas fraccionarias (como lo son los quarks) y ponen en pie de igualdad como constituyentes los bosones y los fermiones. Esto nos hace creer, con Dirac, que "bajo estas circunstancias sería sorprendente que la Naturaleza no hubiese hecho uso de ellos".

La diferenciación química de las células nerviosas

Para el desarrollo de la compleja red de células nerviosas que constituye el sistema nervioso, es necesario que cada célula "elija" una sustancia transmisora que sea la adecuada para los contactos específicos que establece con otras

Paul H. Patterson, David D. Potter y Edwin J. Furshpan

El flujo de información que atraviesa el sistema nervioso viene determinado por el modo de distribución de las conexiones de las neuronas, o células nerviosas, entre sí o con otros tejidos corporales, como los músculos y las glándulas. El modo en que esta red de gran complejidad se establece durante el desarrollo del organismo es una cuestión fundamental de sumo interés. En el embrión, las neuronas, que en un principio están separadas, se ponen en contacto en parte por migración y en parte por formación de finas prolongaciones del cuerpo celular denominadas axones y dendritas. Estas fibras suelen ramificarse repetidas veces y pueden alcanzar zonas del sistema nervioso relativamente distantes o abandonarlo para inervar tejidos efectores. El lugar de contacto entre dos neuronas o entre una neurona y una célula efectora es la sinapsis. En esta zona, la parte terminal del axón aferente llega hasta una proximidad extrema de la célula diana pero sin llegar a tocarla; entre ambas células hay un espacio lleno de líquido que se llama hendidura sináptica. La llegada de un impulso nervioso provoca la liberación de moléculas de una sustancia neurotransmisora que abandonan el pie terminal, atraviesan la hendidura sináptica y se combinan con las proteínas receptoras de gran especificidad que se hallan en la superficie de la célula diana. La activación de los receptores por el neurotransmisor desencadena una respuesta eléctrica o bioquímica.

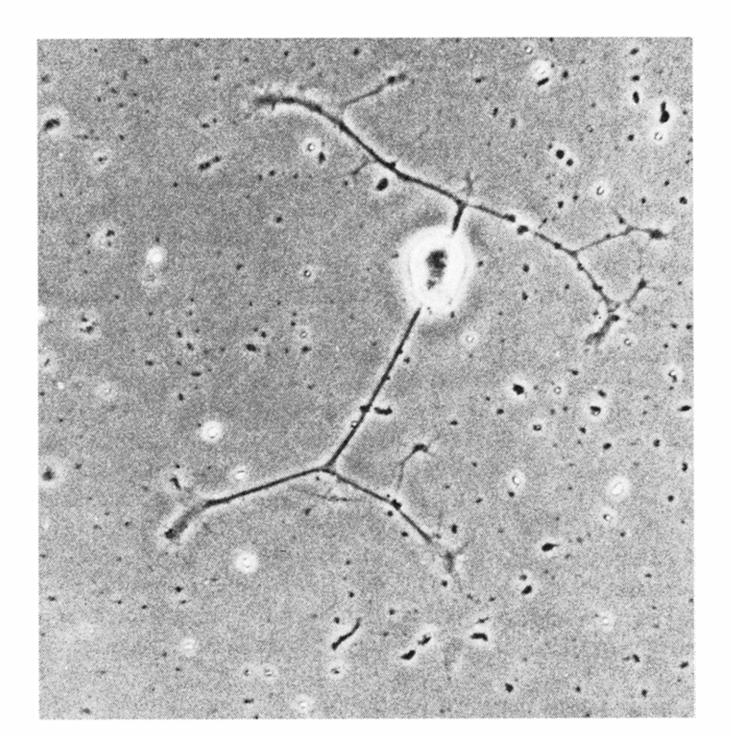
El hecho de que en la arquitectura del sistema nervioso intervengan tanto la química como la topología le da al sistema un orden de complejidad superior al sistema de circuitos creado por el hombre en los computadores. Actualmente, hay más de una docena de sustancias diversas a las que se atribuye el papel de neurotransmisoras, y pueden ser tanto excitadoras o activadoras (aumentando la actividad de la célula diana) como inhibidoras (impidiendo que la célula diana se torne activa). Pero, ¿cómo pueden las neuronas, durante el desarrollo del sistema nervioso, "decidir" cuál es el neurotransmisor que han de segregar de acuerdo con su situación y su función? y ¿cómo se forma la sinapsis emparejando adecuadamente el neurotransmisor segregado por la neurona aferente con los receptores presentes en la superficie de la célula diana? Las investigaciones en torno a estas cuestiones han supuesto la colaboración de electrofisiólogos, bioquímicos y especialistas en microscopía electrónica en una serie de laboratorios.

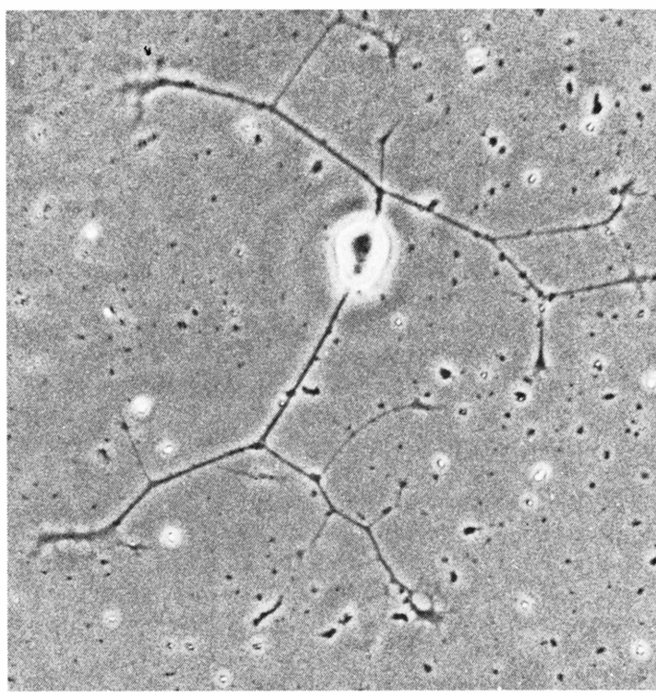
Concretamente en el nuestro, de la Facultad de Medicina de Harvard, con la ayuda de otros 14 colaboradores hemos enfocado el problema general de la diferenciación química de las neuronas estudiando el desarrollo del sistema nervioso autónomo, encargado de regular las actividades del sistema circulatorio y de una serie de órganos y glándulas. La ventaja de dicho sistema en este tipo de trabajo es que las porciones periféricas son de fácil acceso para el investigador. Además, la organización química del sistema nervioso autónomo es relativamente sencilla: la mayoría de las neuronas que la componen segregan bien acetilcolina, bien norepinefrina (noradrenalina); dichas células se llaman colinérgicas y adrenérgicas respectivamente. Parece razonable suponer que el mecanismo básico que entra en juego a la hora de decidir si una neurona autonóma se convertirá en colinérgica o adrenérgica es análogo a los que determinan cuál de varios neurotransmisores posibles será segregado por una célula nerviosa o neurona cerebral.

Encontramos dos tipos de vías principales en el sistema nervioso autónomo: las simpáticas y las parasimpáticas. Estos dos subsistemas a menudo tiene acciones antagónicas sobre un mismo órgano. Así, las células musculares del corazón, que se contraen espontánea y rítmicamente, están inervadas tanto por fibras simpáticas como parasimpáticas. En las situaciones de emergencia o tensión, los nervios simpáticos segregan norepinefrina, que hace latir al corazón más rápidamente y con mayor fuerza. En momentos de tranquilidad o inactividad, los nervios parasimpáticos segregan acetilcolina, que hace latir al corazón con mayor lentitud.

Ambas vías autónomas inervadoras del corazón se inician con neuronas situadas en el cerebro o la médula espinal. Estas neuronas centrales no tienen una proyección directa sobre el músculo cardíaco sino que realizan sinapsis colinérgicas excitadoras con neuronas situadas en un ganglio, estación de relevo localizada fuera de la médula espinal; a su vez, las neuronas ganglionares inervan el músculo cardíaco. Por tanto, la vía simpática consiste en una neurona central colinérgica y una neurona ganglionar adrenérgica, ambas excitadoras; la vía parasimpática consiste en una neurona central colinérgica y una neurona ganglionar colinérgica también, siendo excitadora la central e inhibidora la ganglionar. (La capacidad de la acetilcolina para provocar excitación en la primera sinapsis e inhibición en la segunda se debe a diferencias de tipo químico en las células diana.) Está claro que la regulación normal del corazón depende de la capacidad de las neuronas de ambas vías autónomas para segregar el neurotransmisor adecuado. Por ejemplo, si las neuronas ganglionares simpáticas segregan acetilcolina en lugar de norepinefrina, la inhibición del corazón sería inadecuada.

¿Qué es lo que determina que una neurona específica, situada en una vía autónoma, se convierta en colinérgica o adrenérgica? Para estudiar esta cuestión, hemos llevado a cabo, con nuestros colegas, experiencias con neuronas juveniles



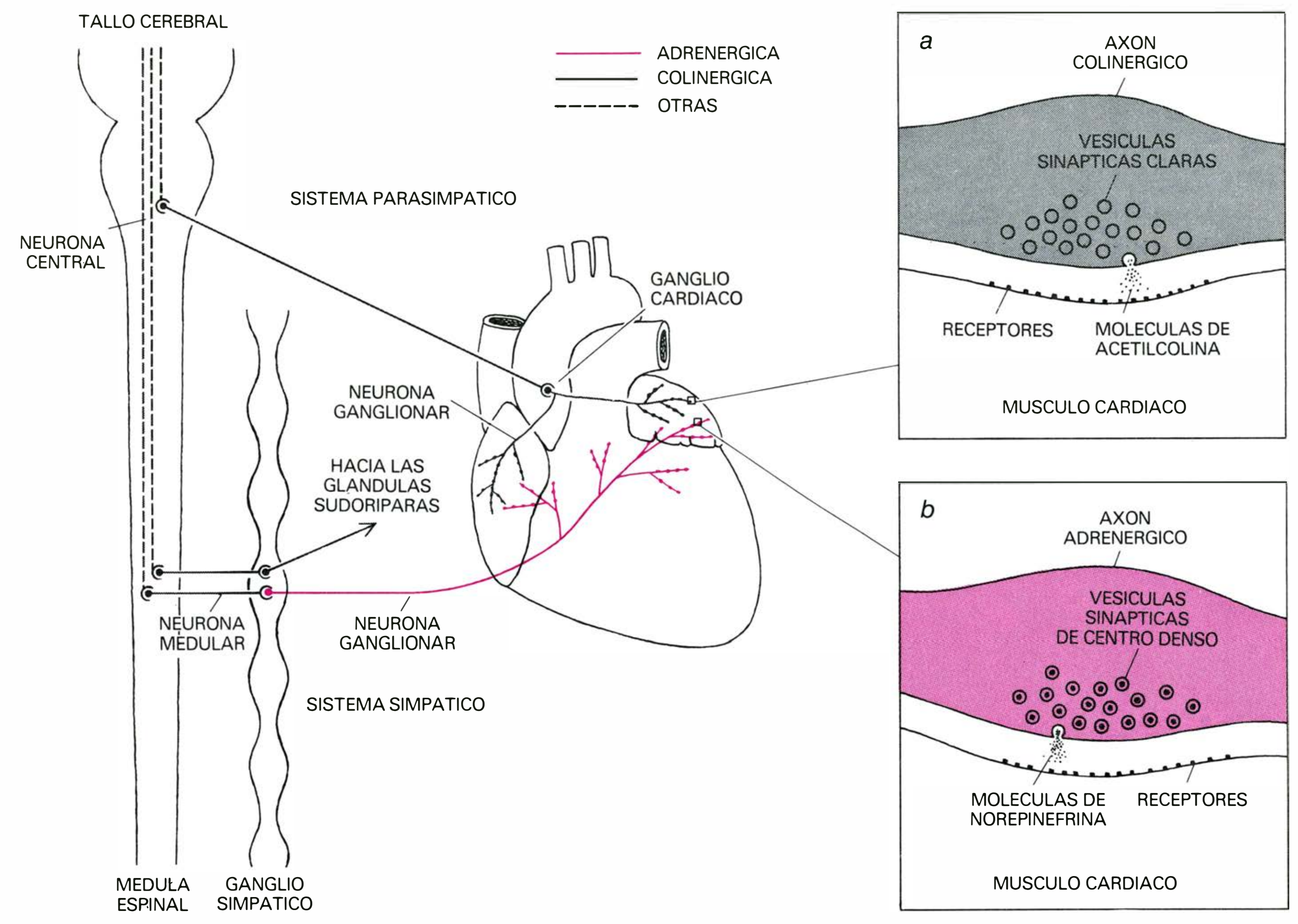


ESTA NEURONA CULTIVADA procede del ganglio simpático de una rata recién nacida; cuando se incuba en presencia de una proteína específica, denominada factor de crecimiento neuronal (NGF), forma unas finas prolongaciones que se van ramificando. La microelectrografía superior izquierda nos muestra la célula pasados 240 minutos de incubación. La superior derecha es a los 300 minutos y la inferior a los 500 minutos. Las prolongaciones crecen de forma continua y, al final, establecen un retículo en la superficie de la placa de cultivo. También pueden formar sinapsis –puntos de comunicación– con otras neuronas o con células musculares cultivadas. La ventaja del estudio de las neuronas en cultivos celulares estriba en que su medio puede ser rigurosamente controlado. Las microelectrografías fueron realizadas por Dennis Bray, de la Unidad de Biofísica Celular del Consejo de Investigación Médica de Londres.

extirpadas de ganglios simpáticos de ratas recién nacidas y posteriormente mantenidas en cultivo de tejidos. Hay tres razones por las que se escogió este tipo de preparado. En primer lugar, los ganglios simpáticos adultos contienen tanto neuronas adrenérgicas como colinérgicas, aunque éstas en número reducido (inervan algunos vasos sanguíneos y las glándulas sudoríparas), de suerte que puede estudiarse la diferenciación de ambos tipos de neuronas en un mismo sistema. En segundo lugar, tal y como lo han demostrado Rita Levi-Montalcini y sus colaboradores de la Universidad de Washington, el desarrollo normal de las neuronas simpáticas en el animal intacto depende de la presencia de una proteína particular, denominada factor de crecimiento neuronal (NGF, de "nerve-growth factor"). Cuando se añade NGF al medio de cultivo, se consigue que las neuronas simpáticas crezcan en el propio cultivo.

Más tarde, Dennis Bray, que se halla actualmente en Londres en la Unidad de Biofísica Celular del Consejo de Investigación Médica, llegó a descubrir unas condiciones de cultivo específicas en las que sobrevivían las neuronas simpáticas de la rata recién nacida pero no así las células no neuronales del mismo ganglio, que morían. Entonces resulta que las neuronas simpáticas se desarrollan en ausencia total de cualquier otro tipo celular. En tercer lugar, cuando se extirpan las neuronas simpáticas de los ganglios de la rata recién nacida, no son todavía maduras en muchos aspectos: en su mayoría, no han llegado a establecer contactos sinápticos y se hallan en una etapa temprana de la diferenciación del metabolismo neurotransmisor.

La ventaja que presenta el estudio del desarrollo de las neuronas en cultivo con respecto a su estudio en el animal intacto es que el investigador tiene un control mayor sobre el medio celular y líquido que rodea a las neuronas y, por tanto, puede estudiar con mayor facilidad los factores esenciales para la inducción o la regulación de funciones celulares determinadas. A las neuronas cultivadas hay que proporcionarles, por supuesto, los nutrientes que normalmente hubieran recibido de la sangre, como pueden ser aminoácidos, azúcares y vitaminas. Un equilibrio apropiado de oxígeno y dióxido de carbono, que normalmente está regulado por el intercambio gaseoso en los pulmones, se puede conseguir por un intercambio de gases entre el medio de cultivo y el ambiente controlado del incubador de cultivos. La eliminación de desechos tóxicos, que normalmente asume el riñón, se realiza con un cambio periódico del medio de cultivo. Al carecer las células aisladas de todo tipo de defensas inmunológicas contra los microorganismos, es necesario que todo el material, tanto instrumentos como placas y medios, sea estéril. Finalmente, al objeto de aumentar la adherencia de los axones y dendritas al sustrato, la superficie plástica de la placa de cultivo se cubre con una fina película de colágeno, que es



DOS VIAS del sistema nervioso autónomo, de acciones antagónicas, controlan la frecuencia de latido del corazón. Cuando el cuerpo se encuentra en una situación de tensión o emergencia y el corazón late más rápido, es que está activo el sistema simpático. Unas vías de reflejo excitan unas neuronas simpáticas que se hallan en la médula espinal. Cuando las neuronas ganglionares se excitan por la llegada de impulsos de la médula, provocan un latido más veloz del corazón mediante la secreción de norepinefrina en el contacto con las células musculares cardiacas. En cambio, cuando el cuerpo está relajado y el corazón late más lento, está activo el sistema parasimpático: las vías reflejo excitan unas neuronas parasimpáticas de la médula, que proyectan sus axones hasta neuronas diana situadas en el ganglio cardíaco. Las neuronas ganglionares, a su vez, hacen que los latidos del corazón sean más lentos, segregando acetilcolina en el contacto con las células miocárdicas. La secreción del neurotransmisor tiene lugar en unas varicosidades: unos pequeños abultamientos colocados a intervalos en los axones segregadores de norepinefrina (adrenérgicos) o de acetilcolina (colinérgicos) que se pueden observar en los cuadros dibujados a la derecha.

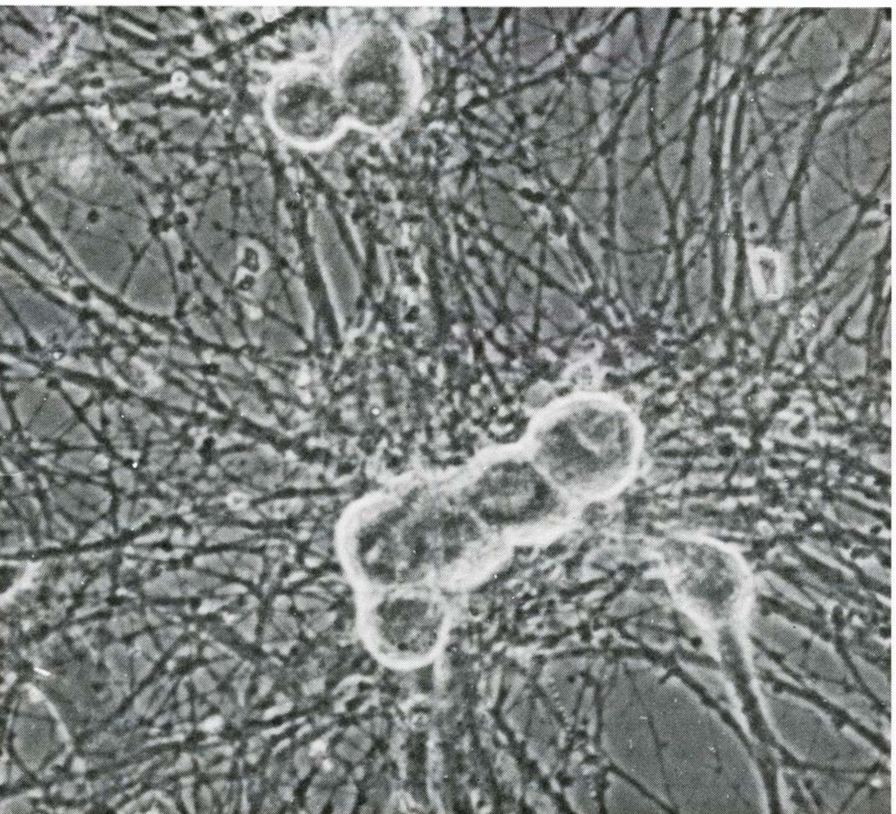
un constituyente natural de la matriz intercelular.

En presencia de NGF, pero no en su ausencia, las neuronas simpáticas sobreviven y forman prolongaciones que se siguen alargando durante muchas semanas. Llega un momento en que la superficie de la placa está toda ella cubierta por un denso retículo de axones y dendritas. Sorprendentemente, las neuronas cultivadas en este medio sumamente artificial adquieren, en ausencia de células no neuronales, la gran mayoría de las características de las neuronas simpáticas normales, evolución que transcurre durante un tiempo similar al observado para dichos cambios en el animal intacto. Las neuronas cultivadas generan impulsos, poseen receptores de superficie para la acetilcolina (el transmisor liberado por las neuronas centrales que inervan normalmente a estas neuronas simpáticas) y desarrollan la capacidad de sintetizar norepinefrina a partir de su precursor aminoacídico, la tirosina. Efectivamente, antes del nacimiento de la rata, la mayoría o quizá la totalidad de las neuronas de los ganglios simpáticos fabrican norepinefrina. Tales observaciones hacen pensar que, antes de ser extirpadas del animal, las neuronas habían recibido señales relacionadas con el desarrollo, dándoles instrucciones para la utilización de norepinefrina como transmisor. En presencia de NGF, factor que permite la supervivencia de las células en cultivo, lo único que hacen las neuronas es leer el programa genético adecuado.

Cuando se hallan en ausencia de células no neuronales, las neuronas simpáticas cultivadas también forman sinapsis entre sí. Como lo han demostrado Richard P. Bunge y sus colaboradores, de la Facultad de Medicina de la Universidad de Washington, estas sinapsis tienen una estructura semejante a las formadas por neuronas adrenérgicas adultas, por llevar norepinefrina almacenada en unas diminutas vesículas, llamadas también sacos, que tienen centros densos cuando se tiñen con permanganato para su observación al microscopio electrónico. Aunque en estas sinapsis se produzca liberación y recepción de norepinefrina, el neurotransmisor no tiene ningún efecto detectable sobre la actividad eléctrica de la neurona diana, lo cual significa o bien que las neuronas cultivadas no poseen receptores para la norepinefrina, o bien que, si están presentes, no consiguen provocar una señal eléctrica significativa. La existencia de estas sinapsis silenciosas es poco común y se conocen escasos ejemplos de este tipo sináptico.

Cuando las neuronas simpáticas se cultivan junto con células no neuronales procedentes del mismo ganglio, se observan diferencias marcadas en la diferenciación

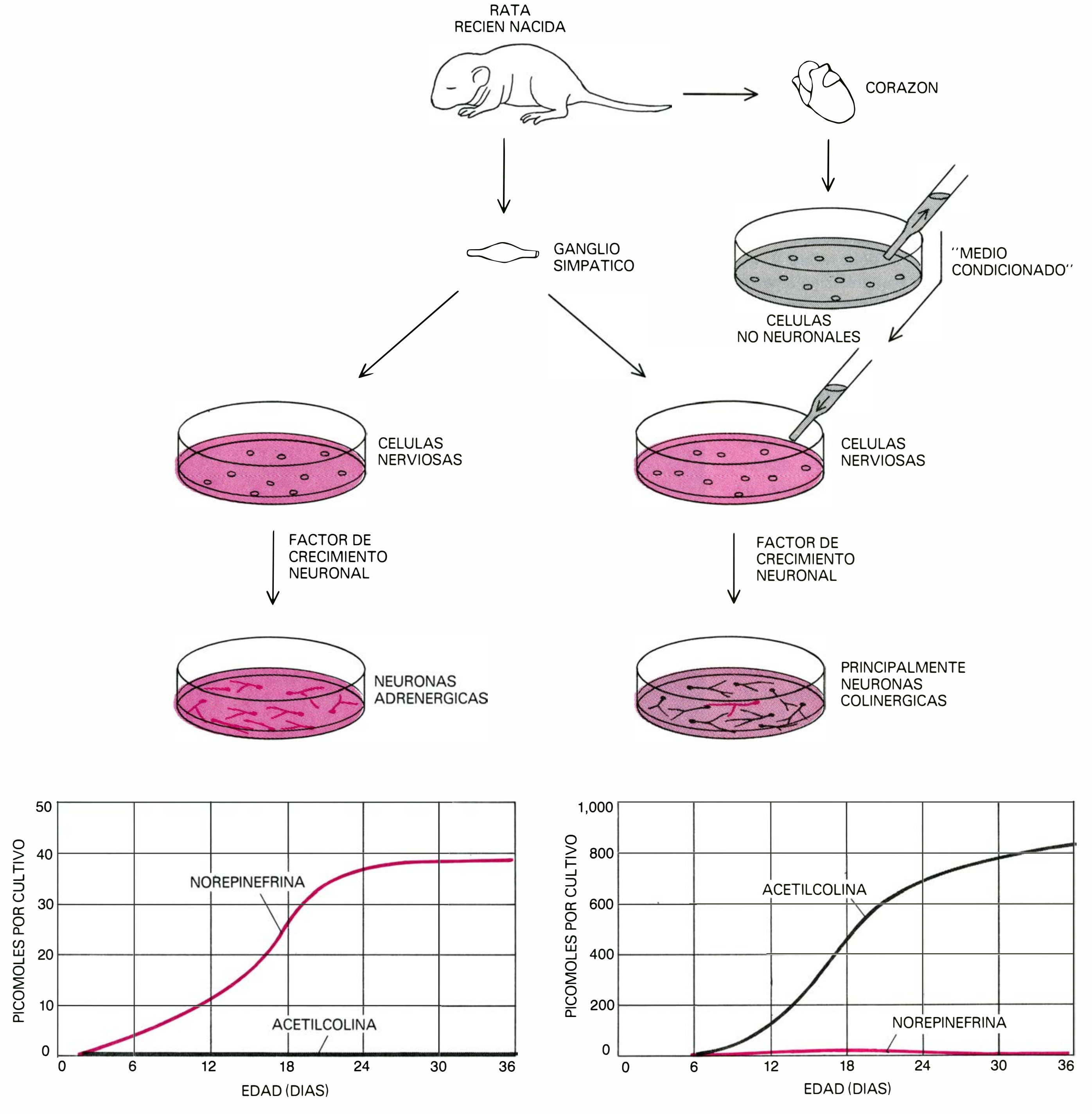




NEURONAS JUVENILES extirpadas de ganglios simpáticos de una rata recién nacida; pueden cultivarse en placas, como se desprende de estas micrografías de cultivos vivos realizadas con microscopio de contraste de fases. En un medio típico, las células no neuronales que actúan de soporte del ganglio también sobreviven y se dividen llegando a formar una capa continua alrededor y por debajo de las neuronas *(arriba)*. Se pueden conseguir cultivos que contengan exclusivamente neuronas *(abajo)* si se cultivan las células ganglionares en un medio en que las neuronas sobreviven, pero no así las células no neuronales. Uno de estos medios posee un compuesto análogo a un precursor del ácido nucleico: arabinósido de citosina, que resulta tóxico al ser incorporado a las moléculas de ácido nucleico de las células no neuronales en división; no afecta, sin embargo, a las neuronas, o células nerviosas que no se dividen sino que simplemente aumentan de tamaño.

química de las neuronas. En vez de manifestar tan sólo funciones adrenérgicas como es el caso en ausencia de células no neuronales, los cultivos mixtos elaboran y acumulan elevados niveles de acetilcolina, sustancia cuya presencia indica el paso de muchas neuronas a colinérgicas. Los estudios electrofisiológicos han puesto de manifiesto que una fracción sustancial de las neuronas también forman sinapsis colinérgicas con sus vecinas, sustituyendo de esta forma la normal recepción de impulsos colinérgicos procedentes de neuronas centrales, las cuales están ausentes del cultivo. A diferencia de las sinapsis adrenérgicas silenciosas desde el punto de vista eléctrico y formadas en ausencia de células no neuronales, estas sinapsis colinérgicas son eléctricamente activas y excitadoras. Uno se puede preguntar si el efecto de las células no neuronales sobre la elección del transmisor se ejerce por medio del contacto directo con las neuronas simpáticas o más bien por medio de la liberación de un factor químico al medio de cultivo.

Para resolver esta cuestión, hicimos cultivos separados de neuronas y de célu-

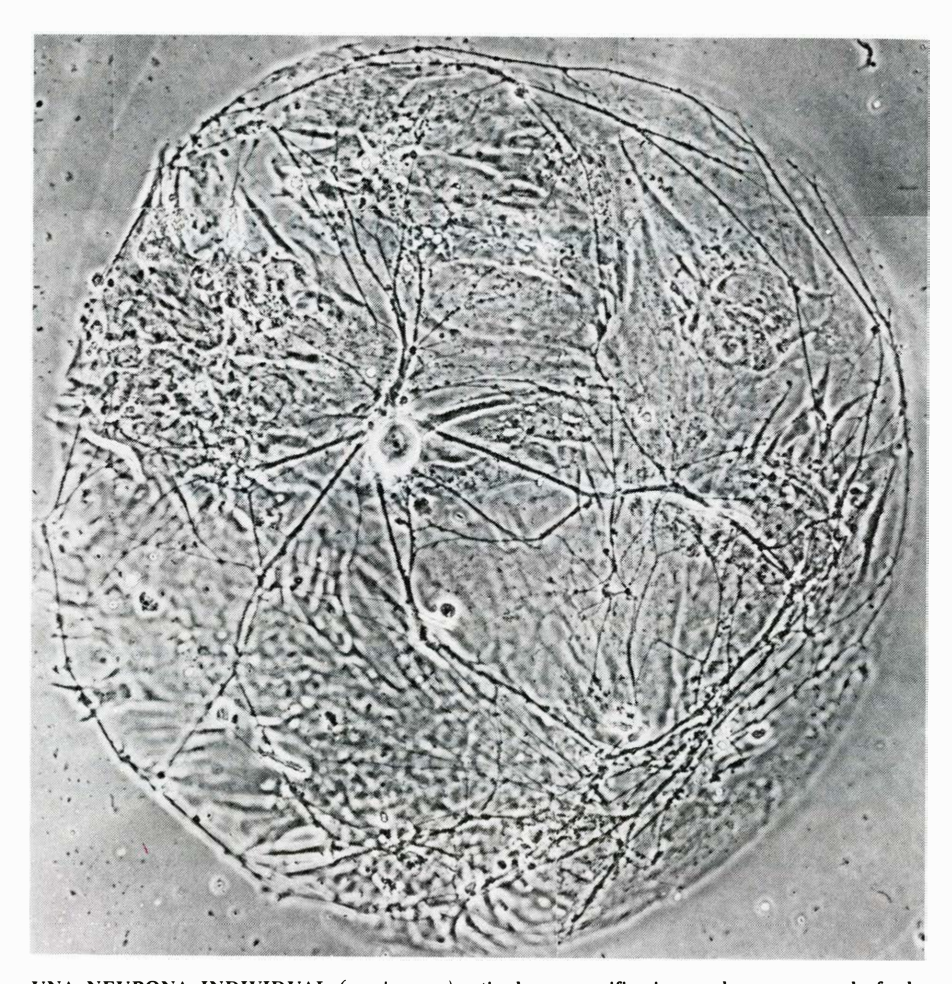


EFECTO DEL MEDIO DE CULTIVO sobre la diferenciación química de las neuronas simpáticas. Se trata de algo sorprendente. Si se extirpan neuronas juveniles de ganglios simpáticos de ratas recién nacidas, y se hacen crecer en cultivos puros, casi todas ellas elaboran y segregan norepinefrina. Sin embargo, si las neuronas simpáticas juveniles se cultivan junto con células no neuronales, o si se tratan con el medio de cultivo que ha sido "condicionado" por incubación con las células no neuronales, la gran mayoría de las neuronas fabrican y segregan acetilcolina. Las gráficas inferiores muestran la expresión de la función colinérgica o adrenérgica en presencia y en ausencia del medio condicionado. Los resultados están expresados en picomoles (10^{-12} moles) de transmisor sintetizado y acumulado en el cultivo de neuronas durante un período de incubación de cuatro horas. En ausencia de medio condicionado *(a la izquierda)*, la capacidad para fabricar norepinefrina aumenta a lo largo de un período de tres semanas y la capacidad para elaborar acetilcolina permanece despreciable. En contrapartida, en cultivos hermanos con un medio que es condicionado en un 62 por cien *(a la derecha)*, la capacidad de síntesis de norepinefrina se eleva en un principio, pero luego disminuye mientras la de acetilcolina aumenta. Por tanto, en presencia de medio condicionado, la expresión de ambos es recíproca.

las no neuronales, en dos series de placas diferentes. Cada dos días se procedía a transferir el medio que había sido "condicionado" por incubación con las células no neuronales a las placas de cultivo de neuronas. Hallamos que el tratamiento con el medio "condicionado" era suficiente para inducir propiedades colinérgicas en las neuronas simpáticas: las células nerviosas sintetizaban y acumulaban acetilcolina y la segregaban en sinapsis colinérgicas funcionales formadas entre estas células.

El efecto del medio condicionado era dependiente de la dosis: cuanto mayor era la cantidad que se añadía al medio de cultivo de las neuronas, mayor era la cantidad de acetilcolina fabricada y mayor también la probabilidad de formación de sinapsis colinérgicas entre estas células. Por otro lado, y concomitantemente, la formación de norepinefrina por parte de las neuronas era menor y se establecían sinapsis adrenérgicas en un número que también era menor.

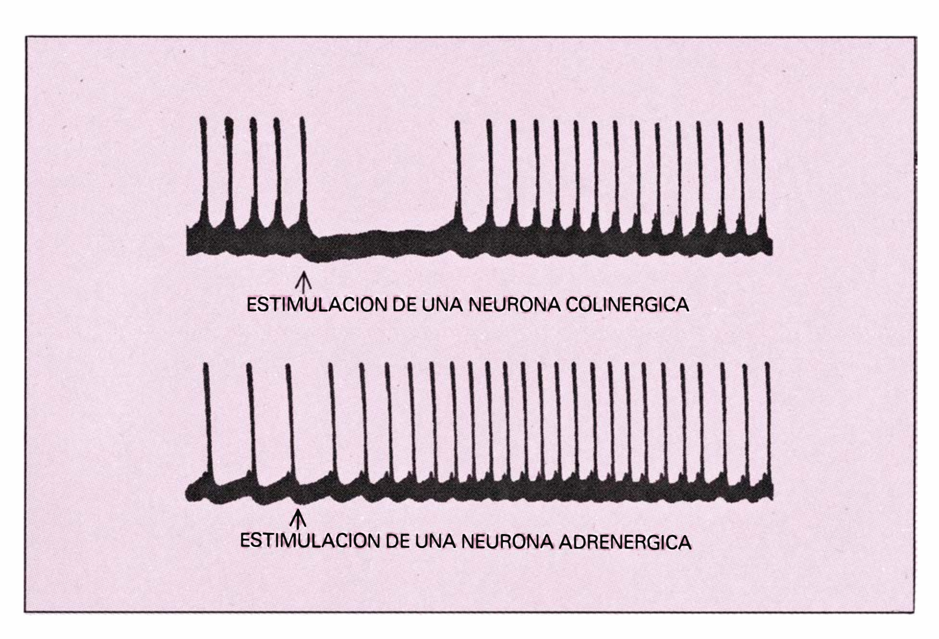
Por tanto, tomando al conjunto de la población de neuronas, la expresión de funciones colinérgicas era aproximadamente recíproca a la expresión de funciones adrenérgicas. Lo que es más, la diferenciación del metabolismo colinérgico en presencia del medio condicionado transcurría durante un tiempo similar al de la diferenciación del metabolismo adrenérgico en ausencia de medio condicionado. Recientemente hemos conseguido pruebas de que el ingrediente activo del medio condicionado es una molécula de gran tamaño, y actualmente se están llevando a cabo intentos para lograr su purificación. de desarrollo liberado por células no neuronales en relación con la elección del neurotransmisor? Una de las posibilidades es que en un principio hubieran dos poblaciones de neuronas, una predestinada a convertirse en adrenérgica y la otra en colinérgica; en ese caso, el factor condicionante favorecería la supervivencia y crecimiento de la población colinérgica disminuyendo concomitantemente la de la población adrenérgica. Esta hipótesis no parece muy probable ya que el medio condicionado no tiene ningún efecto sobre la supervivencia ni sobre el crecimiento de las neuronas cultivadas. Una segunda posibilidad es que el medio condicionado induzca la expresión de propiedades colinérgicas en una población colinérgica predeterminada que, de otro modo, sobreviviría pero sin expresar ningún tipo de metabolismo neurotransmisor. También se desechó esta hipótesis al observar que, en ausencia de medio condicionado, prácticamente ninguna de las neuronas era "silenciosa" con respecto a la síntesis de transmisor.



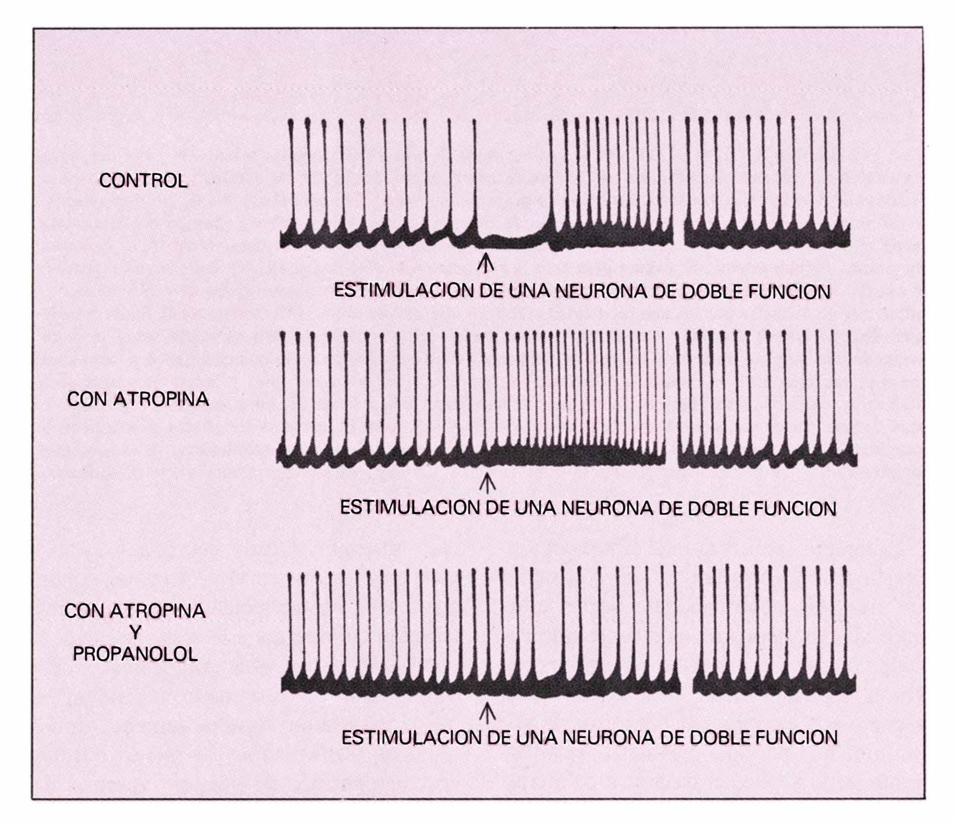
UNA NEURONA INDIVIDUAL *(en el centro)* **extiende sus ramificaciones sobre una capa de fondo constituida por células cardíacas, en un sistema de microcultivo ideado por los autores y sus colegas en la Facultad de Medicina de Harvard. Este tipo de preparación es muy útil para el estudio de la diferenciación de neuronas simpáticas con respecto a la elección de neurotransmisor: el transmisor segregado por la neurona puede ser identificado por su efecto sobre las células musculares cardíacas que, aún en cultivo se contraen espontánea y rítmicamente. El primer paso para la preparación de este microcultivo es hacer una suspensión de células miocárdicas de rata recién nacida. Las células se cultivan sobre unos discos de colágeno de 0,5 milímetros de diámetro por los que las células muestran afinidad de unión. (Mostramos en la figura uno de estos discos.) Uno o dos días más tarde, cuando las células cardíacas se han aplanado sobre el disco formando una capa continua, se siembran las neuronas a baja densidad de modo que en algunos discos habrá una sola neurona. Una vez unida la neurona, ésta empieza a formar unas finas y largas prolongaciones estableciendo conexiones sinápticas con algunas de las células miocárdicas. El neurotransmisor segregado en estas sinapsis puede ser acetilcolina, en cuyo caso se hará más lento el latido de las células miocárdicas, o norepinefrina, en cuyo caso se acelerará. Algunas de las neuronas segregan acetilcolina, otras segregan norepinefrina y, en determinado estadio la diferenciación, las hay que segregar ambos neurotransmisores.**

La tercera explicación del efecto del medio condicionado es la que más concuerda con nuestras observaciones. Según estas hipótesis, las neuronas que expresan propiedades adrenérgicas en un principio conservan todavía cierta "flexibilidad" con respecto a la elección del neurotransmisor durante un período postnatal relativamente largo y pueden pasarse a colinérgicas bajo la influencia del medio condicionado. Este concepto implica que el ingrediente activo del medio condicionado determina la elección del transmisor y el tipo de sinapsis formadas por una neurona simpática, sin afectar por ello a su probabilidad de supervivencia ni a la longitud de su axón y dendritas.

La reciprocidad en la expresión de las funciones adrenérgicas y colinérgicas se puede atribuir a algun tipo de "competencia" inherente a cada neurona y establecida entre la instrucción prenatal conducente a la adrenergicidad y la nueva instrucción conducente a la colinergicidad. Si realmente se da esta competencia, cabe preguntarse si una neurona individual puede manifestar simultáneamente ambos sistemas transmisores, al menos durante un corto período de tiempo. Nuestros intentos por resolver esta cuestión consistieron en cultivar neuronas individuales en un pequeño lecho de células miocárdicas, durante unas dos semanas. Se procedía a introducir microelectrodos en la neurona y en una célula miocárdica adyacente de dichos microcultivos con el fin de tener un registro de su actividad eléctrica; de este modo, se podía determinar la elección del transmisor de la neurona provocando la liberación del transmisor y observando



LA FRECUENCIA DE CONTRACCION de las células miocárdicas de un microcultivo que contiene una sola neurona revela la identidad del neurotransmisor segregado por la neurona. En un experimento típico, un microelectrodo introducido en una neurona estimula la liberación del transmisor, mientras otro microelectrodo introducido en una célula miocárdica inervada registra las contracciones celulares. El trazo superior muestra el efecto de la estimulación de una neurona colinérgica: las contracciones espontáneas de la célula miocárdica cesan temporalmente. El trazo inferior de la ilustración muestra el efecto que producen la estimulación de una neurona adrenérgica: la frecuencia de contracción de la célula miocárdica aumenta.

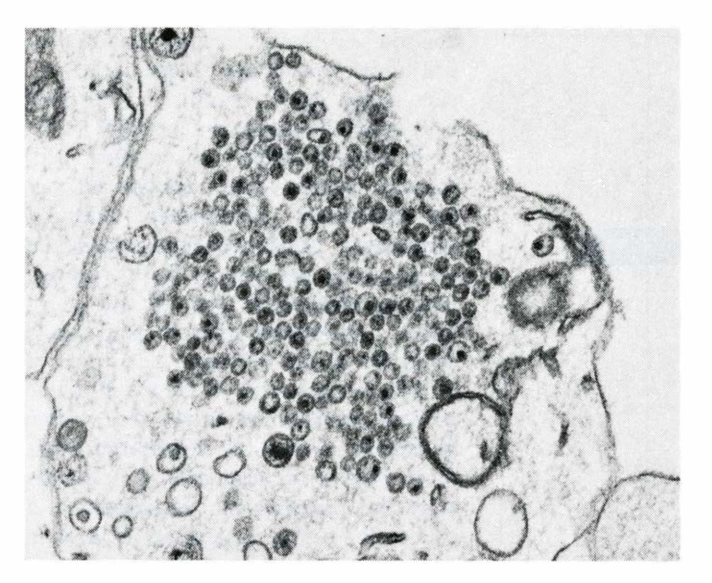


NEURONA DE DOBLE FUNCION presente en un microcultivo de trece días. Nos pone de manifiesto una combinación de propiedades colinérgicas y adrenérgicas. En el trazo superior, la estimulación de la neurona provoca inicialmente un paro en las contracciones de las células miocárdicas inervadas (por mediación de la acetilcolina) seguida por la reanudación de las contracciones a un ritmo acelerado (por mediación de la norepinefrina). Parece que ambos transmisores son segregados a la vez; la separación temporal de la inhibición y la excitación podría deberse al hecho de que los efectos de la norepinefrina sobre las células cardíacas tardan más en aparecer y persisten durante más largo tiempo que los de la acetilcolina. En los dos trazos inferiores, se han añadido al cultivo fármacos que bloquean específicamente bien la transmisión colinérgica bien la adrenérgica, con el fin de verificar la existencia de esta doble funcionalidad. El trazo del centro muestra el efecto de la atropina, que compite con la acetilcolina por su receptor; la inhibición desaparece pero la excitación ulterior permanece inalterada. El trazo inferior de la ilustración muestra el efecto que produce la adición de propanolol (sustancia que compite con la norepinefrina por su receptor) a la solución de atropina: en este último caso quedan bloqueados tanto el efecto inhibidor como el excitador.

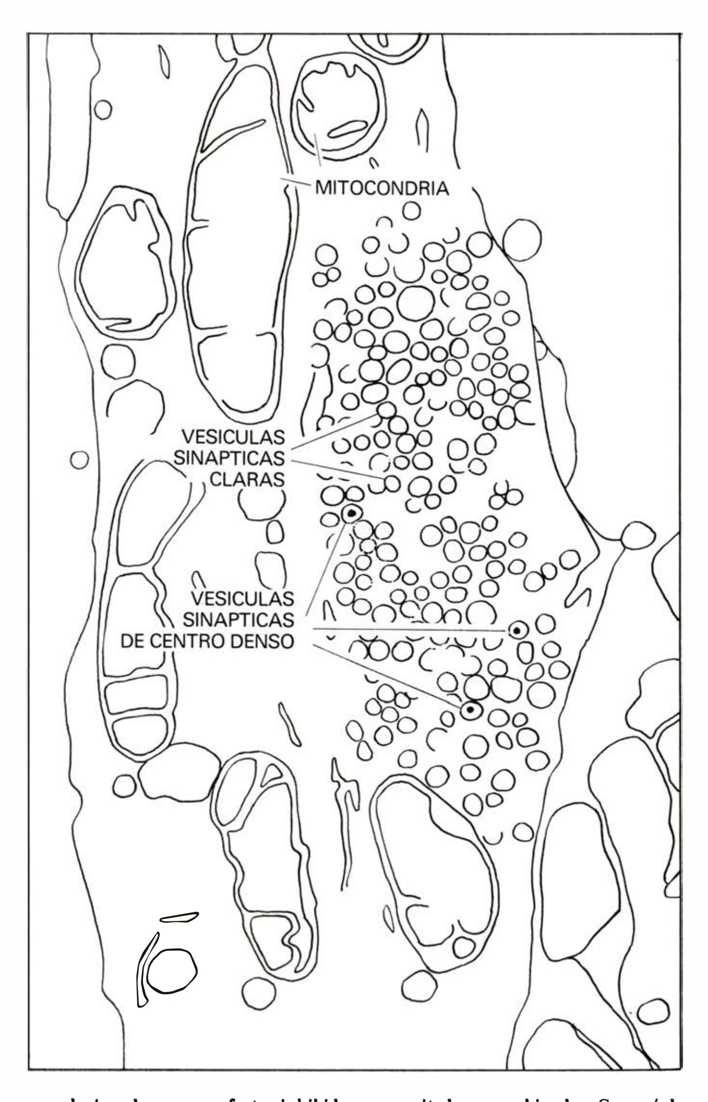
su efecto sobre las células cardíacas, las cuales "laten" espontáneamente y de forma rítmica aún en cultivo. A diferencia de las neuronas simpáticas cultivadas, las células cardíacas poseen tanto receptores colinérgicos como receptores adrenérgicos eléctricamente activos; así, pues, una disminución de la frecuencia (o paro total) del latido indica la secreción de acetilcolina, mientras que la aceleración del latido es señal de liberación de norepinefrina. Se pueden completar las pruebas para la identificación del transmisor, observando los efectos producidos por determinados fármacos, que compiten específicamente con el transmisor natural, competencia centrada en los lugares de unión de los receptores de las células miocárdicas. Así, la atropina bloquea los receptores cardíacos de acetilcolina y el propanolol bloquea los receptores cardíacos de norepinefrina (a veces llamada noradrenalina).

La utilización de esta técnica en nuestro trabajo nos ha llevado a identificar tres tipos de neuronas en los microcultivos de dos semanas. El primer tipo es adrenérgico: excita a las células cardíacas y dicho efecto queda bloqueado por el propanolol. El segundo tipo es colinérgico: inhiba las células cardíacas y el efecto queda bloqueado por la atropina. El tercer tipo de neurona manifiesta tanto actividad adrenérgica como colinérgica: la estimulación de la neurona inhibe al corazón por un mecanismo sensible a la atropina pero, posteriormente, lo acelera mediante un mecanismo sensible al propanolol. Ya que en cada microcultivo se cuenta con la presencia de una sola neurona, parece evidente que la misma célula es mediadora de ambos efectos. Cuando se examinan los microcultivos al microscopio electrónico, se observa un gran número de vesículas sinápticas con el centro denso en el interior de la neuronas adrenérgicas, vesículas claras en las neuronas colinérgicas y unas cuantas vesículas de centro denso combinadas con un número mucho mayor de vesículas claras en las células de doble función. Una ventaja evidente del estudio de estos cultivos es que contienen tan sólo una neurona, por lo que se puede establecer correlaciones carentes de ambigüedades entre su estructura y su función.

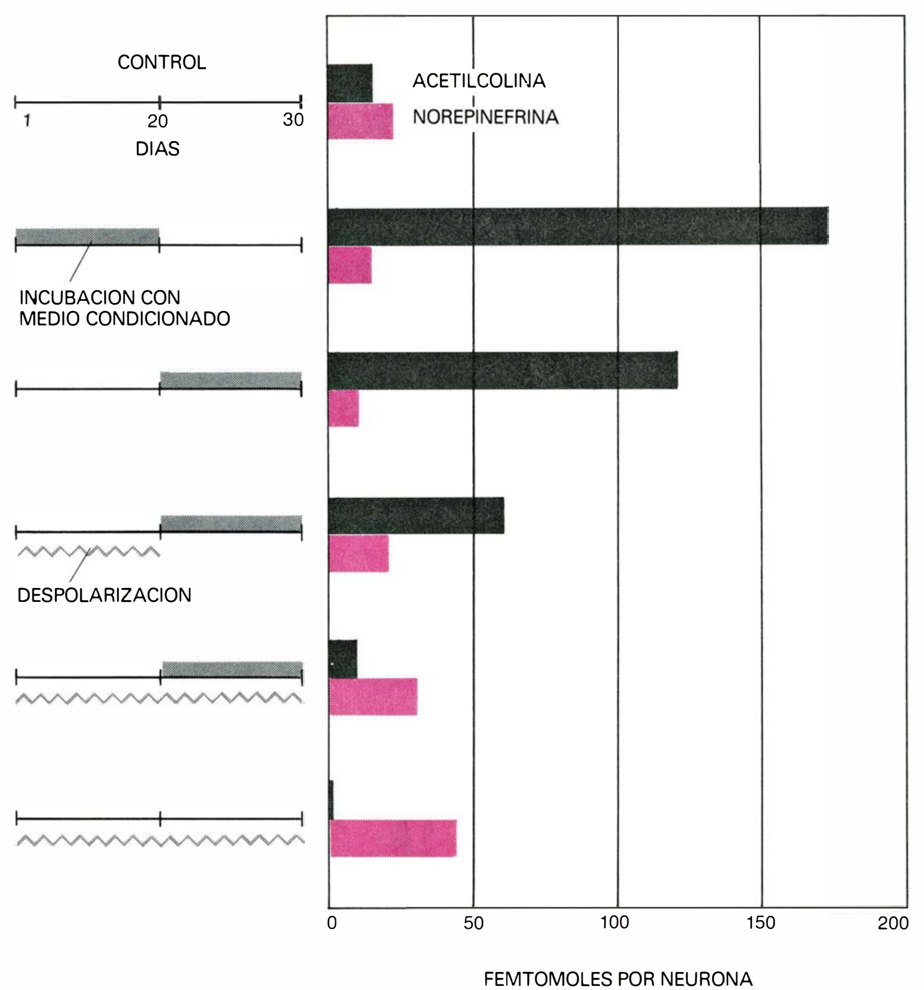
Estos resultados nos indican que una sola neurona es capaz de expresar simultáneamente ambos sistemas de transmisores, cuando aún se halla en un estadio temprano. La doble función aparece como un concepto nuevo, pero en anteriores investigaciones nunca se habían estudiado neuronas en proceso de desarrollo de un modo que permitiera poner en evidencia este bicomportamiento (o comportamiento dual). Pero, pensándolo bien a la vista de




LAS VARICOSIDADES SINAPTICAS que se hallan a lo largo de los axones de neuronas cultivadas, colinérgicas o adrenérgicas, evidencian características diferentes en su ultraestructura. Las moléculas de neurotransmisor están almacenadas en unos sacos diminutos llamados vesículas sinápticas. Cuando las varicosidades de las neuronas adrenérgicas se tiñen con permanganato y se observan al microscopio electrónico, se ve que contienen vesículas sinápticas con centros pequeños y densos a los electrones *(a la izquierda).* Por otra parte, las viscosidades de las neuronas colinérgicas tienen vesículas de interior claro *(a la derecha).* Estas dos microelectrografías, al igual que la de la figura inferior fueron realizadas por Story Landis, de Harvard, con 44.000 aumentos.

VARICOSIDAD DE DOBLE FUNCION, que contiene unas cuantas vesículas de centro denso, de las características de neuronas adrenérgicas *(véase el mapa)* junto con una gran mayoría de vesículas claras, de las características de neuronas colinérgicas. Por tanto, la ultraestructura fina de la varicosidad está correlacionada con sus efectos inhibidores y excitadores combinados. Se pasó de una neurona adrenérgica a una colinérgica bajo la influencia de medio condicionado. La duración de la doble función es un dato que todavía se desconoce. No hay superposición temporal en la síntesis de las enzimas que intervienen.



EFECTO DE LA ACTIVIDAD ELECTRICA sobre la diferenciación química de las neuronas simpáticas. Este consiste en reducir su capacidad para convertirse en colinérgicas al ser sometidas a medio condicionado. En el ganglio intacto, la actividad es el resultado de la inervación excitadora procedente de neuronas de la médula espinal. La influencia excitadora de las neuronas medulares se puede simular en cultivos celulares tratando las neuronas simpáticas con elevadas concentraciones del ion potasio, el cual despolariza las neuronas (invierte el potencial o voltaje que atraviesa la membrana). En el experimento que aquí se representa, se añadía potasio o medio condicionado en los días 0 o 20, y se determinaba cada transmisor elaborado por las neuronas en el día 30. (Un femtomol equivale a 10^{-15} moles). La despolarización de las células con potasio bloqueaba en gran parte el efecto del medio condicionado. Por consiguiente la pequeña fracción de neuronas del ganglio simpático intacto que terminan siendo colinérgicas quizás no posean actividad eléctrica propia hasta después de ser expuestas a inducción por parte de las células no neuronales.

las pruebas aportadas, la doble función aparece como un paso intermedio lógico en la conversión de una neurona adrenérgica en colinérgica bajo la influencia del medio condicionado. Aun teniendo en cuenta que no existe superposición temporal en la síntesis de los enzimas y demás componentes que participan en la elaboración y liberación de ambos neurotransmisores, no parece descabellado suponer que los enzimas y las vesículas sinápticas mediadores de la transmisión adrenérgica sigan funcionando durante algún tiempo después del cese de su síntesis. Todavía no se conoce la duración exacta de este estadio de doble función.

Pasado un período de aproximadamente cuatro semanas, las neuronas individuales de los microcultivos se hacen tan grandes que se les puede someter a ensayos bioquímicos. Se incuban las células en una mezcla de tirosina y colina radiactivas durante unas ocho o doce horas; luego se determina la cantidad de norepinefrina y de acetilcolina que ha sido sintetizada a partir de las moléculas precursoras. En ausencia de medio acondicionado, prácticamente todas las neuronas sintetizan cantidades detectables tan sólo de norepinefrina. En presencia de células cardíacas, por el contrario, una mayor sustancia de las neuronas fabrican sólo acetilcolina. En ninguna de las condiciones estudiadas se observa un número significativo de neuronas "silenciosas" ni de neuronas de doble función, pasadas cuatro semanas de cultivo.

Estas observaciones indican que la mayoría de las neuronas son adrenérgicas en el momento de ponerlas en cultivo, pero que son sensibles a un mecanismo de control transformador que determina la elección última del transmisor. La duración del período de transición en el que se expresa la doble función no se conoce, pero pasadas cuatro o cinco semanas después del nacimiento, prácticamente todas las neuronas se han diferenciado, alcanzando un estado en que sólo se expresa de manera significativa un único sistema transmisor. Actualmente, estamos siguiendo la evolución de las neuronas individuales a lo largo de un período de tiempo mediante pruebas intermitentes de sus efectos sobre las células miocárdicas. Esperamos llegar así a observar directamente la transición entre el comportamiento adrenérgico y el colinérgico y poder determinar la sucesión de eventos que la caracterizan.

¿En que momento de la vida de cultivo es máxima la sensibilidad de las neuronas respecto a la acción del medio condicionado, y hasta qué punto son reversibles los efectos de éste? Buscamos la respuesta añadiendo un medio condicionado a cultivos de neuronas simpáticas en una estimulación continuada de diez días de duración, repitiendo la misma operación en cultivos de diversos estadios de madurez celular. Resultó que la capacidad de respuesta de las células a la estimulación del medio condicionado disminuía rápidamente al aumentar la edad de las células, alcanzando un nivel bajísimo cuando la estimulación se aplicaba entre los 40 y 50 días. Caryl Hill y Ian Hendry, de la Universidad Nacional Australiana, y C. David Ross, Mary Johnson y Bunge, de la Facultad de Medicina de Harvard, obtuvieron resultados similares en animales intactos. Si se toman pequeños trozos de ganglios simpáticos de ratas de edades diversas y se colocan en cultivo durante un período típico, la expresión de la función colinérgica declina progresivamente al aumentar la edad del animal al que se le extirparon los trozos, siendo dicha función minimizada en los ganglios de las ratas adultas.

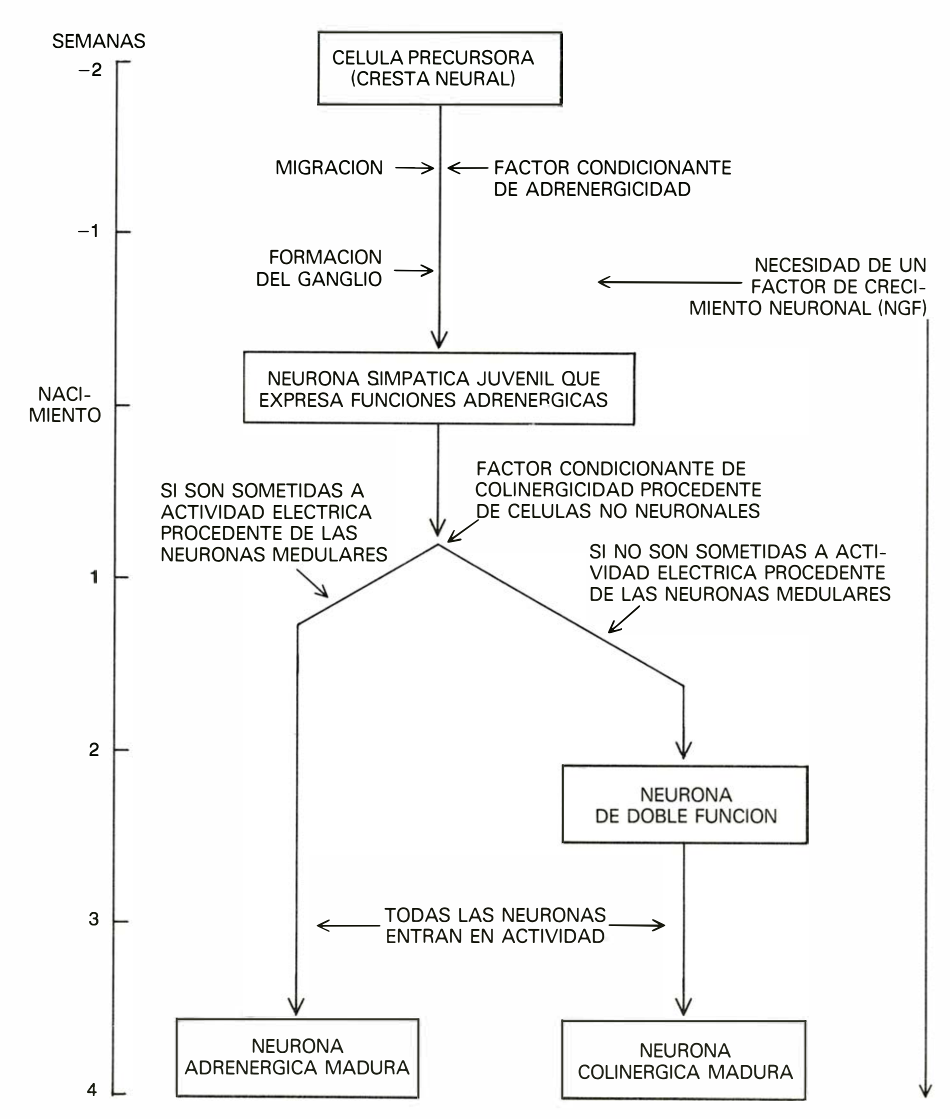
Uno de los aspectos que más intriga despierta, en relación con la influencia de las células no neuronales sobre la elección del neurotransmisor, es la distribución de la eficacia de dicho efecto entre las diversas células no neuronales del cuerpo. Si esta propiedad estuviera muy localizada, podría actuar como un determinante específico de la elección del transmisor. Nuestra aproximación al asunto fue estudiar el efecto de varios tejidos procedentes de la rata recién nacida sobre dicha elección. Todas las células no neuronales eran capaces de inducir funciones colinérgicas, pero existían diferencias cuantitativas claras en cuanto a la eficacia

de la acción: las células del músculo esquelético eran las más eficaces, las del corazón tenían una eficacia intermedia y las hepáticas eran las menos activas. Un rasgo evidente de este orden de eficacias en su relación con la cuantía de la inervación colinérgica recibida por cada uno de los tejidos diana: las fibras musculares esqueléticas reciben exclusivamente inervación colinérgica las células hepáticas, si es que reciben inervación alguna, es de tipo adrenérgico. Esta observación sugiere que, durante el desarrollo, las células diana liberan factores químicos que influyen en el metabolismo del transmisor de las neuronas que inervan a dichas células diana.

Sin embargo, otras observaciones no parecían quedar enmarcadas en esta hipótesis sencilla. Por ejemplo, había varios tipos de células que eran eficaces como condicionadoras del medio, pero se trataba de células que normalmente no están inervadas, en el sentido de que no reciben sinapsis. Entre ellas tenemos a los fibroblastos (células del tejido conectivo), a un tipo de células tumorales de la glía que se han propagado en serie mediante cultivos (las células de la glía son las células de soporte de las neuronas en el cerebro y en la médula espinal) y a las células no neuronales de los ganglios simpáticos. Parecía particularmente paradójica la capacidad de las células ganglionares no neuronales para inducir la diferenciación colinérgica. Durante las tres primeras semanas después del nacimiento, las células no neuronales del ganglión rodean de cerca a cada neurona; a pesar de ello, la mayoría de las neuronas del ganglio intacto terminan siendo adrenérgicas. ¿Por qué razón las células no neuronales o el medio condicionado son capaces de hacer que casi todas las neuronas simpáticas se vuelvan colinérgicas en los cultivos y, en cambio, en el organismo tan sólo pocas neuronas simpáticas pasen a colinérgicas incluso en presencia copiosa de células no neuronales?

Esta aparente paradoja indicaba que, en la elección del transmisor, intervenían nuevas variables. Las neuronas simpáticas del cuerpo estaban expuestas a la señal colinérgica que proporcionan las células no neuronales, aunque algún factor que faltaba en el medio de cultivo evitaba que la mayoría de ellas respondiese a la señal. Es posible que este factor fuese el NGF, ya que es imprescindible para la supervivencia y crecimiento de las neuronas simpáticas adrenérgicas, tanto en el animal intacto como en cultivo. Es más, si se expone a los cultivos neuronales a altas dosis de NGF, hay un aumento de producción de componentes específicos adrenérgicos, como norepinefrina, en relación con la producción de componentes celulares no específicos, tales como las proteínas y los lípidos. Sin embargo, resulta que el NGF tiene el mismo efecto potenciador sobre la diferenciación colinérgica. Por tanto, con respecto a la producción de transmisor, el NGF es más permisivo que instructivo: estimula el crecimiento y diferenciación de neuronas simpáticas inmaduras tanto por vía adrenérgica como colinérgica, pero no influye en cuál de ellas se tome. A diferencia del NGF, el factor colinérgico secretado por células no neuronales no afecta a la supervivencia ni al crecimiento neuronal, pero sí instruye a las neuronas con respecto a la elección del neurotransmisor.

Otro candidato para el factor presente en el animal intacto y cuya presencia impedía el paso a colinérgicas de la mayoría de las neuronas simpáticas era la inervación excitadora normal procedente de las neuronas centrales, las cuales estaban ausentes de nuestros cultivos. Por entonces, Ira Black, Hendry y Leslie Iversen, trabajando en el Laboratorio de Biología Molecular del Consejo de Investigación Médica, en Cambridge (Inglaterra), habían comprobado que si se corta la llegada de impulsos de la médula espinal a los ganglios simpáticos de ratones de poca edad, se aminora el desarrollo ulterior del metabolismo adrenérgico. Estas observaciones hicieron surgir la posibilidad de que la actividad eléctrica a la que están sometidas las neuronas de los ganglios durante la primera semana después del nacimiento desempeña un papel decisivo en determinar la elección del transmisor. Aunque las neuronas centrales inervadoras no están presentes en los cultivos celulares, conseguimos simular su efecto excitador sobre las neuronas ganglionares (efecto que consiste en la despolarización de la membrana de la neurona ganglionar, es decir, en la



ESTE DIAGRAMA DE FLUJO representa una hipótesis relativa a los pasos del desarrollo implicados en la diferenciación química de una neurona simpática. Según esta hipótesis, las influencias ambientales que afectan a la elección del neurotransmisor por parte de la célula se ejercen tanto por células no neurales (que liberan factores de desarrollo) como por otras neuronas (cuya actividad eléctrica modifica la respuesta de la neurona a los factores anteriormente mencionados). El eje del tiempo es aproximado. Ya que el desarrollo de las neuronas no es sincronizado, cada una de ellas atraviesa la vía de desarrollo a tiempos diferentes.

inversión del potencial que existe de un lado al otro de la membrana). La simulación se hizo elevando la concentración de potasio del medio, añadiéndole el fármaco veratridina (que provoca la entrada de iones sodio a las neuronas) o estimulando eléctricamente a las células con una frecuencia de un estímulo por segundo y con una duración de varios días.

Cuando se despolarizaban cultivos masivos de neuronas, bien en presencia de medio condicionado, bien durante siete o diez días antes de la adición de medio condicionado, las neuronas permanecían primariamente adrenérgicas. Efectivamente, la despolarización hacía bajar la relación entre la síntesis de acetilcolina y la síntesis de norepinefrina hasta llegar a valores 300 veces menores que los de cultivos que recibían simplemente medio condicionado. Los cambios mencionados se daban sin una alteración significativa en la supervivencia de las neuronas, lo que sugiere que las neuronas, que en presencia de medio condicionado se hubieran convertido en colinérgicas, ahora se mantenían adrenérgicas. Era como si la despolarización de las neuronas simpáticas hubieran estabilizado su instrucción prenatal de adrenergicidad y reducido en gran parte su flexibilidad con respecto a la elección del transmisor.

La actividad eléctrica de las neuronas va acompañada de una entrada de iones calcio en su interior. Basándose en el hecho de que la entrada de iones calcio es importante para el control de la secreción de neurotransmisores y probablemente de otras funciones celulares, se planteó la cuestión de si los efectos de la despolarización sobre la elección del transmisor se podrían reducir impidiendo la entrada de calcio. Esta situación se llevó a cabo elevando la concentración de iones magnesio en el medio, o añadiendole el fármaco D600 que bloquea selectivamente el movimiento de los iones calcio a través de la membrana celular. Efectivamente, como se suponía, al bloquear la entrada de calcio, la despolarización ya no impedía la inducción de propiedades colinérgicas por el medio condicionado. En consecuencia, el efecto de la despolarización sobre el bloqueo del efecto del medio condicionado sobre la elección del transmisor parece darse por mediación de los iones calcio, aunque no se conoce todavía su mecanismo exacto.

Todas estas observaciones subrayan la posibilidad de que en los ganglios intactos la mayoría de las neuronas estén mantenidas en su condición adrenérgica prenatal por medio de la llegada de impulsos eléctricos de la médula espinal, llegada que se inicia durante la primera semana después del nacimiento. Una hipótesis-corolario de la anterior sería que las neuronas de los ganglios que luego serán colinérgicas (las que inervan algunos vasos sanguíneos y las glándulas sudoríparas) tan sólo recibieran impulsos eléctricos después de haber sido sometidas a la influencia de las células no neuronales. Existe una posibilidad interesante inherente en dicho mecanismo. Quizá la formación selectiva de sinapsis entre las neuronas centrales y las neuronas ganglionares no sólo establezca el sistema de circuitos de las vías del sistema nervioso autónomo sino también determine la elección del transmisor apropiado para dicho sistema de circuitos. Este concepto ensalza la importancia potencial de la actividad neuronal de la diferenciación química del sistema nervioso.

Los experimentos descritos en este artículo no dejan lugar a dudas respecto a que las neuronas simpáticas en proceso de desarrollo son flexibles, al menos de forma pasajera, en relación con la elección del transmisor. También han llegado recientemente a una conclusión similar los integrantes del grupo de Nicole Le Douarin, del Instituro de Embriología de Nogent-sur-Marne en Francia, que llevan a cabo experimentos ingeniosos de trasplante de embriones de ave. En su opinión, puede alterarse la decisión de las neuronas juveniles en cuanto a su transmisor trasplantando las células a lugares del embrión distintos del normal. Por ejemplo, una población precursora que de suyo origina neuronas colinérgicas dará neuronas adrenérgicas si se trasplanta a una región del organismo donde normalmente surgen neuronas adrenérgicas. En resumen, tanto en cultivos como en los enbriones intactos, esta importante decisión neuronal no está totalmente programada y puede verse influenciada por otras células, tanto si son neuronas como si no.

Lo que aquí se ha debatido ha sido el control de una sola decisión en el desarrollo de un solo tipo de células nerviosas. El desarrollo de la compleja red sináptica del sistema nervioso adulto probablemente implique un gran número de decisiones, muchas de las cuales dependan de interacciones celulares. Con el fin de desenredar estas interacciones, es probable que sea necesario combinar varios tipos de enfoques, como el estudio del desarrollo en el embrión y su paralelo en condiciones de cultivo, simplificadas y controladas. En última instancia, sería satisfactorio llegar a poder especificar cada elemento que hay que añadir a una placa de cultivo para que una neurona embrionaria siga exactamente la misma sucesión de pasos y comportamientos que en el organismo intacto, desde el momento en que cesa la división celular hasta su envejecimiento.

Los rayos cósmicos de máxima energía

Un pequeño número de partículas muy rápidas procedentes del espacio tienen mayor energía que las producidas por los aceleradores. Se manifiestan mediante los chubascos de otras partículas que producen en el aire

John Linsley

En todo momento y procedente de todas las direcciones, está cayendo sobre la Tierra una lluvia continua de partículas que se mueven con velocidad próxima a la de la luz. Estas partículas, llamadas rayos cósmicos, han atraído la atención de los investigadores desde hace más de seis décadas. Si uno estudiara los rayos cósmicos en la superficie de la Luna, donde están ausentes la atmósfera y la influencia del campo magnético terrestre, encontraría que son principalmente núcleos de los elementos más abundantes, es decir, aquellos cuyo peso atómico oscila desde el más ligero, el hidrógeno, hasta el grupo del hierro (peso atómico 56). Además, habría unos pocos núcleos de elementos más pesados, algunos electrones y positrones, unos pocos rayos gamma y los omnipresentes pero escurridizos neutrinos. Estas partículas, por lo demás familiares, se han denominado rayos cósmicos porque se encuentran atravesando el espacio exterior a casi la velocidad de la luz, tan raudamente como las partículas en el haz de un sincrotón u otro acelerador construido por el hombre. De hecho, algunas partículas componentes de los rayos cósmicos tienen una energía mucho más elevada que la de las más rápidas partículas generadas por los mayores aceleradores nunca construidos. Son estos rayos cósmicos altamente energéticos los que plantean más problemas. ¿De dónde vienen? ¿Cómo han sido acelerados hasta tales energías?

Pese al hecho de que los rayos cósmicos altamente energéticos son partículas subatómicas –es decir, los objetos más pequeños que se conocen– sus energías cinéticas son tan grandes que a ese respecto se igualan con los objetos de la vida diaria. Los rayos cósmicos más energéticos entre los ahora observados tenían energías de dos a tres veces 10^{20} electronvolt, lo que es comparable a la de una pelota de tenis muy bien restada.

Los rayos cósmicos en la región de las energías más altas, digamos por encima de 10^{15} electronvolt, tienen un carácter especial porque al tratar de responder preguntas que pertenecen a la astrofísica, quienes trabajan en este campo encuentran continuamente preguntas sin respuesta que competen a un campo de la ciencia ahora llamado física de altas energías. Este ya era el estado de la cuestión en la investigación de los rayos cósmicos para energías por debajo de 10^{15} electronvolt. Se tardó mucho tiempo, tras el vuelo en globo que condujo a Victor F. Hess a su descubrimiento de los rayos cósmicos, en hallar una respuesta a la pregunta básica: ¿Qué son los rayos cósmicos?

Al principio, los físicos trataron de hallar una respuesta usando las mismas técnicas que se acostumbraba a usar para identificar las rápidas partículas emitidas por materias radiactivas o las generadas por los primitivos "destructores de átomos". Tropezaron con dificultades. Encontraron, por ejemplo, que cuando los rayos cósmicos penetraban en la atmósfera, prácticamente nunca llegaban a quedar en reposo poco a poco, que es lo que habrían hecho si sus energías iniciales hubieran sido tan bajas como las de las partículas rápidas que los investigadores estaban acostumbrados a identificar. Por el contrario, cuando siguieron las trayectorias de lo que vinieron a llamarse rayos cósmicos primarios, encontraron que a medida que las partículas penetraban la materia, casi todas las trayectorias conducían a puntos donde había tenido lugar un violento choque entre la partícula del rayo cósmico y un núcleo atómico. De aquellos puntos partían las trayectorias de muchas partículas subatómicas rápidas llamadas rayos cósmicos secundarios.

Los complejos fenómenos asociados a esta clase de colisiones demostraron ser tan interesantes y tener tal importancia para aumentar nuestro conocimiento de la naturaleza, que el estudio de dichos fenómenos en sí mismos prosiguió con creciente vigor. Cuando se dispuso de máquinas capaces de acelerar artificialmente electrones y protones hasta energías tan grandes como las de al menos algunos rayos cósmicos, el estudio de estos fenómenos alcanzó independencia en forma de una nueva rama de la ciencia llamada física de altas energías.

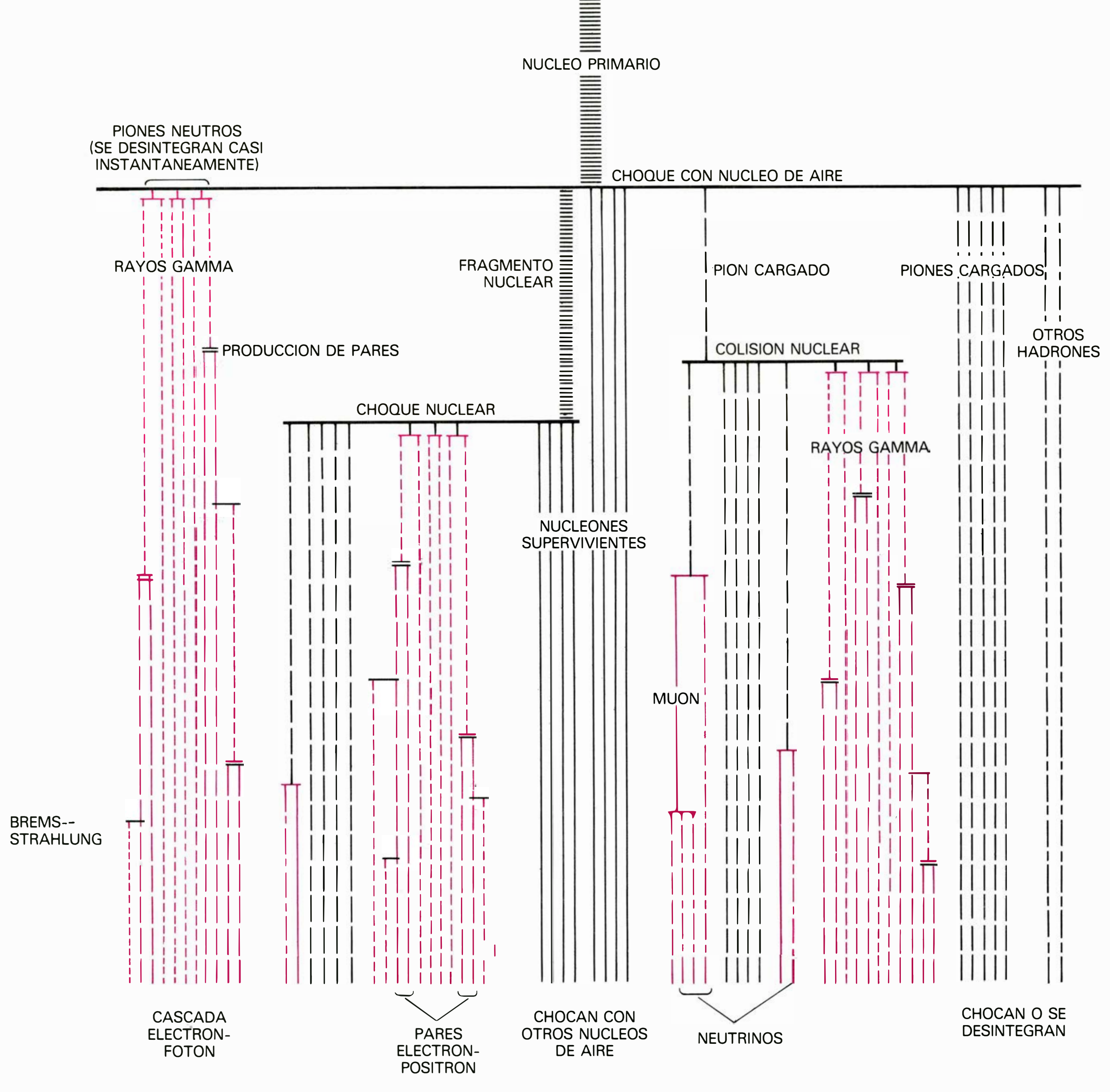
No hay que decir que si es posible realizar un experimento en física de altas energías usando partículas aceleradas artificialmente, es mucho mejor hacerlo así que usando rayos cósmicos. En la actualidad, virtualmente todo lo que sucede en las colisiones entre rayos cósmicos y núcleos que hacen de blanco se puede duplicar en laboratorios de aceleradores hasta energías de rayos cósmicos de unos 10^{12} electronvolt. Se están proyectando aceleradores que en el curso de los próximos diez años serán capaces de elevar esta cifra hasta 10^{15} electronvolt. El coste de su construcción, sin embargo, será muy elevado y las nuevas máquinas bien podrían ser las últimas de su clase.

El hecho de que choques tales como los que se producen entre rayos cósmicos y núcleos atómicos del aire se puedan ahora estudiar usando partículas aceleradas artificialmente ha cambiado el carácter de los experimentos de rayos cósmicos en el dominio de energía por debajo de 10^{12} electronvolt. Los experimentadores sacan provecho de la precisa y detallada información publicada por los físicos de altas energías para calcular la manera como sus propios instrumentos responderán a la radiación cósmica. O bien simplemente transportan sus instrumentos a un laboratorio dotado con acelerador y los calibran colocándolos en un haz de partículas idéntico a la componente de rayos cósmicos que intentan medir. Por otra parte, difícilmente ningún experimento de rayos

cósmicos en el dominio de energías por debajo de 10^{12} electronvolt tiene como objetivo adquirir nueva información acerca de choques de alta energía.

Los experimentos en el dominio de energías entre 10^{12} y 10^{15} electronvolt tienen un carácter intermedio. Por encima de 10^{15} electronvolt, en el dominio de los rayos cósmicos de más alta energía, todavía tienen el carácter que acostumbraban a tener todos los experimentos de rayos cósmicos. Cuando se trabaja en este dominio de energías, uno puede ser, intencionadamente o a la fuerza, astrofísico un día y físico de altas energías el siguiente.

Cualquiera que sea el punto de vista, aparecen en el camino las mismas dos grandes dificultades. Una es que la atmósfera terrestre tiende a impedir que la radiación cósmica primaria alcance el nivel del suelo. La probabilidad de que un protón incidente, por ejemplo, alcance el nivel del mar sin chocar con un núcleo atómico atmosférico es sólo de alrededor de uno en un millón. Los rayos cósmicos secundarios producidos en los choques con los núcleos atmosféricos tienden a ser partículas inestables que se desintegran en otras. Como resultado, la radiación cósmica presente al nivel del mar está formada casi



UN CHUBASCO AEREO se produce cuando un núcleo de un rayo cósmico primario tal como una partícula alfa (núcleo de helio) penetra en la atmósfera terrestre a unos 20 kilómetros sobre el suelo y choca con el núcleo de un átomo en la atmósfera. Este diagrama, altamente esquemático, muestra las primeras fases en la evolución del chubasco. Las líneas en color representan leptones: partículas tales como los fotones (incluyendo rayos gamma), neutrinos, electrones, positrones y muones. Las líneas en negro representan hadrones: partículas más pesadas tales como los nucleones (protones y neutrones), núcleos compuestos y piones. Las partículas que se producen en las dos colisiones secundarias que se muestran constituyen la segunda generación de partículas. Las partículas que ellas producen si chocan con otras partículas antes de su desintegración constituyen la tercera generación, y así sucesivamente. Los hadrones pueden crear tanto leptones como más hadrones, pero los leptones sólo pueden crear más leptones. "Producción de pares" se refiere a la creación de pares electrón-positrón por un fotón. "Bremsstrahlung" (que significa radiación de enfrentamiento) se refiere a fotones creados por la deceleración de otras partículas.

exclusivamente por productos de desintegración: muones, neutrinos y rayos gamma, junto con electrones que son productos secundarios de la formación de muones y rayos gamma. Si un astrofísico trata de estudiar la radiación primaria directamente, tiene que enviar detectores por encima de la atmósfera mediante globos, cohetes o satélites artificiales. El físico de altas energías que pretende estudiar choques entre núcleos de alta energía y núcleos atómicos del blanco puede verse en la necesidad de hacer lo mismo o al menos establecer su laboratorio en una montaña alta.

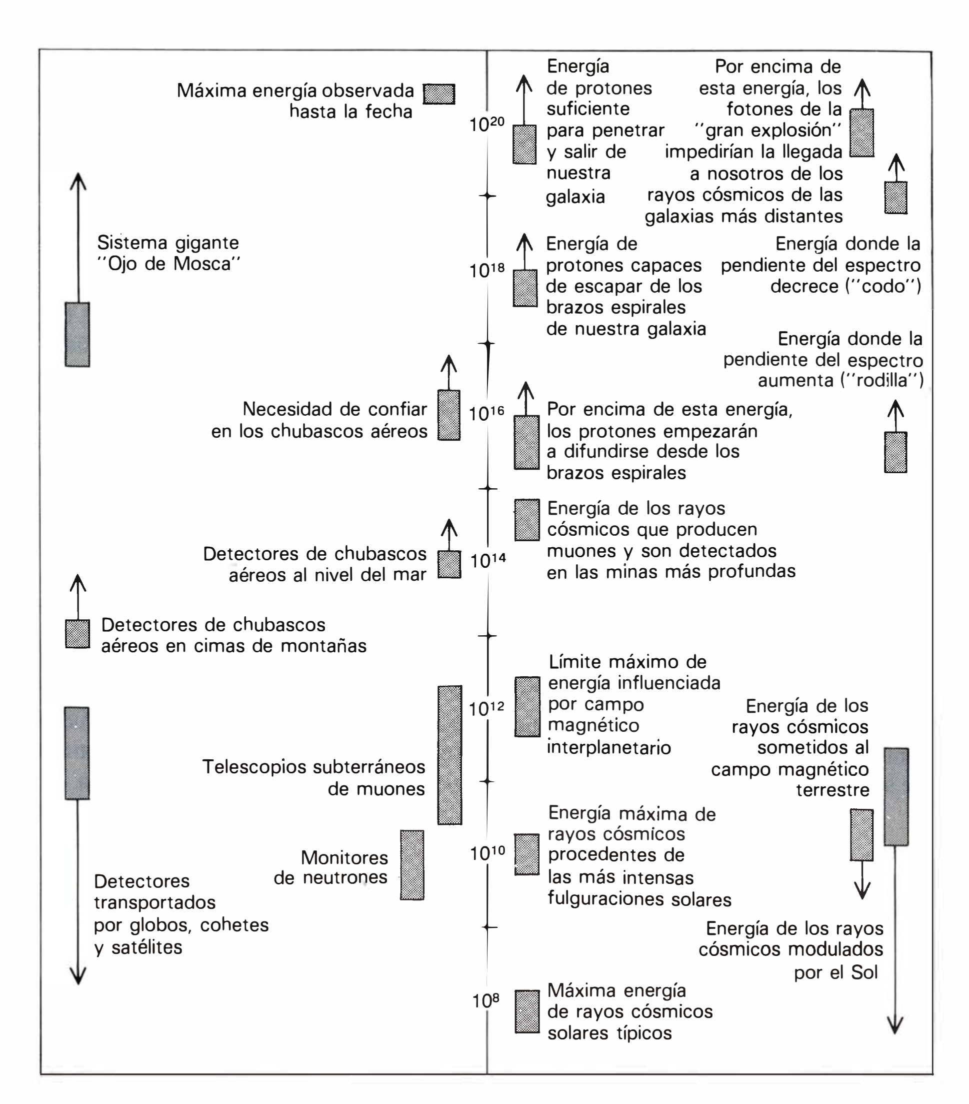
La otra dificultad, todavía mayor, es que la intensidad de la radiación cósmica es muy baja. Aun en la banda de energía más favorable (de 10^9 a 10^{10} electronvolt) y en el lugar más favorable (por encima de la atmósfera terrestre) sólo hay intensidad bastante para producir unos diez choques por minuto en un área de un centímetro cuadrado. Esta cadencia es tan baja que a pesar de la alta energía de las partículas individuales de los rayos cósmicos, la energía total que suministran por unidad de tiempo a tal blanco no es mayor que la recibida en forma de luz de las estrellas. Además, la intensidad de lo que aquí estoy llamando rayos cósmicos de máxima energía es considerablemente más baja. La fracción de los rayos cósmicos primarios con energías superiores a 10^{15} electronvolt es tan pequeña que para capturar una partícula por hora habría que disponer de un dispositivo colector de al menos 10 metros de lado. Por supuesto, independientemente del tamaño, no existe dispositivo capaz de capturar tal partícula con su energía intacta.

¿Cómo pudieron entonces Pierre Auger y sus colegas de la Escuela Normal Superior de París deducir, como lo hicieron en 1938, la existencia de algunos rayos cósmicos primarios con energías tan altas como 10^{15} electronvolt, a partir de los datos obtenidos con unos pocos tubos contadores Geiger que tenían un área total de menos de un metro cuadrado? Lo hicieron poniendo de manifiesto que las descargas simultáneas que observaban en los contadores eran originadas por un fenómeno de rayos cósmicos previamente desconocido y mostrando algunos de los caracteres esenciales del fenómeno. Atribuyeron las descargas simultáneas a chubascos de partículas de rayos cósmicos que incidían en sus instrumentos procedentes del aire situado encima. Partiendo del hecho de que se producían cuentas simultáneas aun cuando lo contadores Geiger estuvieran separados por muchas decenas de metros, dedujeron que los chubascos eran extensos, cubriendo un área de varios cientos de metros cuadrados, y que el número de partículas rápidas en los chubascos era grande, quizá decenas de miles.

Los chubascos descubiertos por Auger y sus colegas se llaman chubascos aéreos extensos o simplemente chubascos aéreos. Sin embargo, el nombre no debe tomarse al pie de la letra. Al contrario de lo que sucede con los chubascos de lluvia o de confetti, los chubascos de rayos cósmicos tienen una estructura muy complicada pero bien definida, y empiezan y acaban en un abrir y cerrar de ojos. Detectando los chubascos aéreos, mejor que las partículas primarias que los producen, uno puede superar las dos dificultades descritas más arriba. El equipo para estudiar los chubascos puede estar situado al nivel del suelo: no necesita transportarse hasta más allá de la atmósfera. Además, el instrumental no necesita cubrir completamente un área grande. Como los chubascos aéreos son extensos, pueden ser detec-



LOS CHUBASCOS AEREOS se detectan mediante un sistema gigante de detectores identificados por los puntos en esta fotografía aérea de Haverah Park en los alrededores de Leeds, Inglaterra. El diámetro del área aproximadamente hexagonal cubierta por los detectores es de unos cuatro kilómetros. Cada detector está constituido por numerosos depósitos herméticamente cerrados que contienen agua muy pura con sensores de luz. Los sensores (tubos fotomultiplicadores) responden a la luz Cerenkov generada por los rayos cósmicos secundarios que penetran en el agua. El área combinada de 32 detectores es de 500 m².



LA ENERGIA DE LOS RAYOS COSMICOS observados abarca un enorme intervalo de 10^8 a 10^{20} electronvolt (*derecha*). Deben tomarse en cuenta tanto el intervalo de energías como las limitaciones de los distintos tipos de detectores cuando se trata de investigar diferentes tipos de rayos cósmicos (*izquierda*). La longitud de los rectángulos, para las partículas y los detectores, indica el intervalo típico de energías. Si las flechas apuntan hacia arriba, el intervalo es un límite inferior; si hacia abajo, un límite superior.

tados a cierta distancia. Se puede sensibilizar un área grande al impacto de los chubascos distribuyendo adecuadamente en ella un conjunto de pequeños detectores de partículas rápidas ampliamente espaciados, y se puede obtener datos a un ritmo razonable a pesar de su muy pequeña intensidad.

El año anterior al descubrimiento de los chubascos aéreos por Auger y sus colegas, se había elaborado un explicación teórica de chubascos análogos de rayos cósmicos pero producidos en blancos sólidos, casi simultáneamente por Homi J. Bhabha y Walter H. Heitler en Inglaterra y por J.F. Carlson y J. Robert Oppenheimer en los Estados Unidos. De acuerdo con esta interpretación, los chubascos aéreos fueron al principio considerados como simples cascadas electrón-fotón iniciadas en la alta atmósfera por electrones de los rayos cósmicos primarios. Existe ahora un gran volumen de pruebas de que casi todos los rayos cósmicos primarios que originan chubascos aéreos son núcleos atómicos, y de que los chubascos de aire son bastante más complicados de lo que al principio se imaginó.

El alma de un chubasco aéreo es una cascada nuclear que lleva consigo una cadena de interacciones nucleares. Comienza con una colisión entre dos núcleos atómicos, uno de ellos la partícula del rayo cósmico primario y otro el núcleo de un átomo de aire. Muchos, pero no todos, los rayos cósmicos secundarios producidos por ese choque sufren a su vez colisiones a lo largo del eje de la cascada, originando sus propios rayos cósmicos secundarios y así sucesivamente. En cada generación de colisiones, sin embargo, algunos de los rayos cósmicos producidos se desintegran, particularmente los piones.

Los piones cargados producidos en las fases preliminares de la cascada generalmente chocan con los núcleos de átomos de aire en vez de desintegrarse, porque sus vidas medias son muy largas cuando se toma en cuenta la dilatación del tiempo según la teoría de la relatividad especial. Los producidos en las últimas fases de las cascadas suelen desintegrarse porque sus energías son más bajas y el efecto relativista no es tan grande. Las partículas que resultan de la desintegración de los piones cargados son muones y neutrinos. Los piones neutros son tan inestables que se desintegran inmediatamente, dando lugar a pares de fotones de alta energía. Estos fotones, que reciben una fracción importante de la energía liberada en cada choque, inician las cascadas electrón-fotón. Brevemente, un chubasco aéreo de rayos cósmicos consta de una cascada de nucleones (protones y neutrones) en su origen, junto con dos grupos de partículas subatómicas rápidas procedentes de procesos de desintegración: una componente muón-neutrino y una componente electrón-fotón.

La componente electrón-fotón está formada por un gran número de cascadas electrón-fotón. Tal cascada lleva también consigo una cadena de choques. Sin embargo, éstos son choques en que un electrón o fotón de alta energía interacciona con el campo eléctrico de un átomo. Las potentes fuerzas de corto alcance que ligan entre sí a los nucleones para formar núcleos atómicos no desempeñan aquí papel alguno. En las fases iniciales de una cascada electrón-fotón, el número de partículas rápidas aumenta porque, por una parte, un fotón de alta energía puede materializarse cuando encuentra un intenso campo eléctrico formando dos partículas, un electrón y un positrón, y por otra parte, un electrón o positrón de alta energía que encuentra un intenso campo eléctrico puede radiar una parte importante de su energía en forma de fotón de alta energía. A medida que el número de partículas rápidas aumenta, sin embargo, la energía media por partícula tiene que decrecer. Con el tiempo, entran en juego ciertos procesos disipativos y la cascada se extingue.

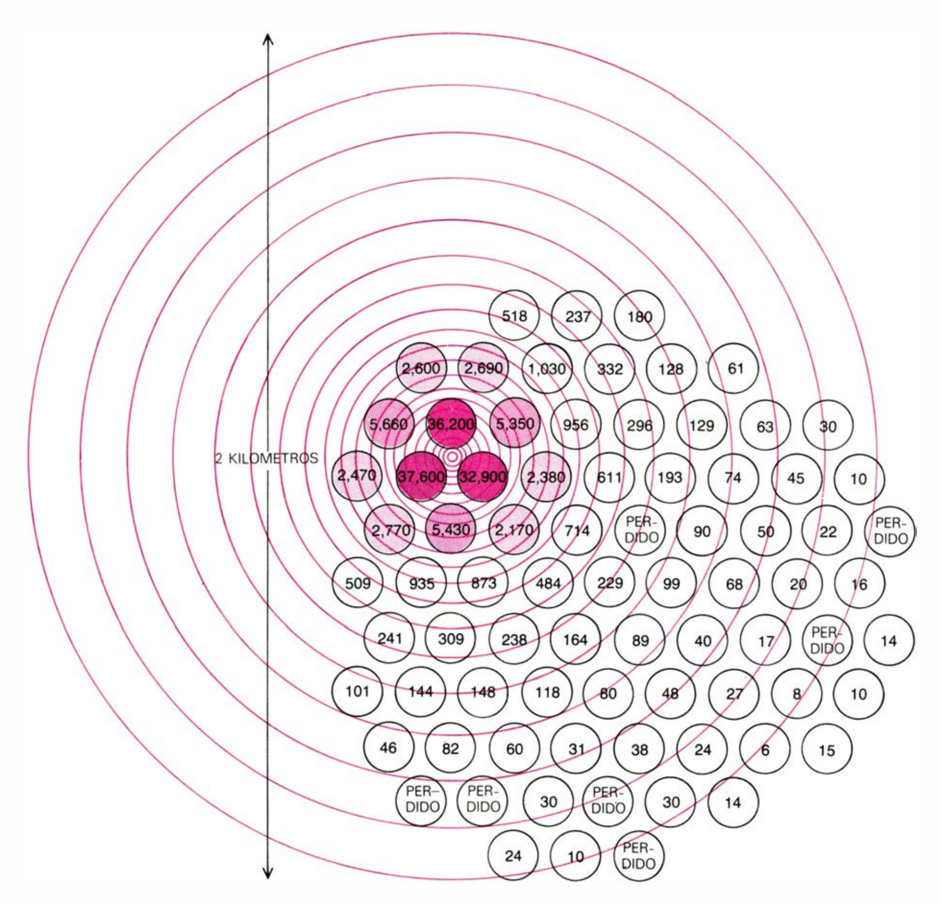
Se puede visualizar un chubasco aéreo construyendo un modelo a escala 1: 20.000, consistente en un disco de medio centímetro de diámetro recortado de papel grueso y enhebrado en un hilo de un metro de largo. El hilo representa el eje del chubasco o prolongación de la trayectoria de la partícula primaria. Si el hilo se mantiene vertical, el extremo superior corresponde a una altura típica de la primera colisión (admitiendo que el extremo inferior está al nivel del mar). Para representar la breve existencia de un chubasco aéreo, uno debe imaginarse el disco desli-

zando hacia abajo a lo largo del hilo a la velocidad de la luz.

El disco representa la región en la que se van a encontrar electrones. Los electrones son, con mucho, las partículas cargadas más abundantes en los chubascos aéreos. La región interna en la cual tiene lugar la cascada de nucleones es solamente del tamaño del agujero para sacar el hilo, unos pocos metros de diámetro en el chubasco real. El disco es una representación bastante buena de la distribución longitudinal de los electrones. En la región central de los chubascos aéreos reales, el espesor vertical de la zona de electrones es sólo de un metro aproximadamente. En ángulo recto con el eje, el disco representa la situación en forma más simple de lo que realmente es.

Contrariamente a lo que ocurre con la densidad del papel, que es constante hasta un cierto radio y entonces cae bruscamente a cero, la densidad de electrones en un chubasco decrece suave y continuamente a distancias crecientes del eje, sin que se observe corte alguno. El disco en el modelo representa la región que contiene la mitad de los electrones. En los chubascos aéreos reales, el diámetro de esta región es de unos 100 metros independientemente de la energía de la partícula que inició el chubasco. La densidad de partículas en el chubasco, sin embargo, aumenta cuando lo hace la energía primaria a todas las distancias desde el eje del chubasco, de modo que aun a distancias del eje tan grandes como un kilómetro, la densidad de partículas en chubascos de muy alta energía es suficientemente grande para que los chubascos se puedan detectar con contadores de tamaño modesto.

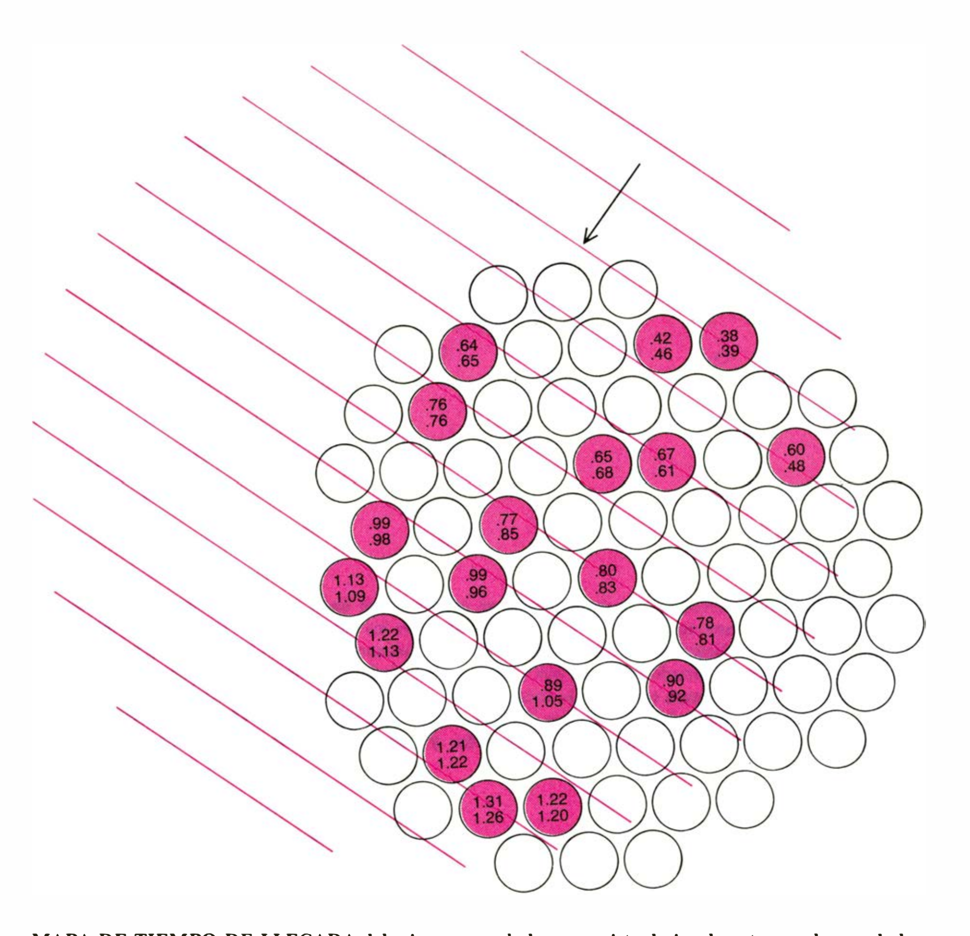
Si se pudiera seguir un chubasco aéreo en la dirección de su movimiento y observar su crecimiento, se vería que el número de partículas aumenta rápidamente al principio, alcanza un máximo allí donde la creación de nuevas partículas está compensada por la absorción y decrece a partir de entonces. El número de partículas en el chubasco en la fase llamada de plenitud o crecimiento máximo, es decir, el máximo número de partículas en un chubasco aéreo dado, es proporcional a la energía de la partícula de rayos cósmicos primarios que generó el chubasco. Por ejemplo, un rayo cósmico primario con una energía de 10^{15} electronvolt generará un chubasco aéreo que contiene medio millón de partículas en su fase de plenitud. Tales chubascos de energía comparativamente baja alcanzan su plenitud en la alta atmósfera. Sin embargo, un chubasco de esa energía todavía puede ser detectado con facilidad al nivel del mar si el rayo cósmico primario se propagase en dirección casi vertical cuando penetró en la atmósfera. Los chubascos aéreos producidos por los rayos cósmicos más energéticos hasta ahora observados, aquellos cuyas energías valen dos o tres veces 10^{20} electronvolt, parecen alcanzar su plenitud cuando han penetrado aproximadamente ocho décimas de atmósfera, es decir, aproximadamente a una altura de 2000 metros sobre el nivel del mar.



MAPA DE INTENSIDAD de un chubasco aéreo muy grande producido por un rayo cósmico de energía 10^{19} electronvolt registrado por un sistema de detectores en Rancho Volcano, Nuevo México, en marzo de 1973. Las líneas concéntricas en color corresponden a igual valor de la intensidad, expresada en función del número de partículas por metro cuadrado. El eje del chubasco era casi vertical, de modo que las líneas de igual intensidad son casi circulares. Los círculos negros representan contadores de centelleo de un metro de diámetro separados entre sí por 150 metros de terreno llano. (Los detectores no están dibujados a escala; si lo hubieran estado, cada uno habría sido más pequeño que el punto al final de esta frase.) Los números dentro de cada detector representan el número de partículas del chubasco medidas en ese punto. Las señales más altas indican que el eje del chubasco pasa cerca del centro del triángulo formado por los detectores.

Los experimentos de Auger y colaboradores abrieron una puerta para estudiar los rayos cósmicos de máxima energía al demostrar que los chubascos aéreos proveen una manera de compensar la rápida disminución de intensidad de los rayos cósmicos cuando su energía aumenta. Sin embargo, en esos experimentos los chubascos se detectaron meramente, sin medirlos individualmente. Es por sí mismo evidente que del hecho aislado de que se produjeran descargas en el mismo momento en tubos de Geiger ampliamente espaciados no se puede determinar el número de rayos cósmicos que intervenían o hallar sus direcciones. No se puede ni siquiera estar seguro de que las descargas coincidentes correspondían a un chubasco aéreo, puesto que puede haber descargas independientes que por casualidad se produzcan al mismo tiempo. Las conclusiones a que llegaron Auger y sus colaboradores, es decir, que existen chubascos aéreos y que algunos de ellos tienen energías de al menos 10^{15} electrónvolt, estaban basadas en el análisis estadístico de muchos sucesos ayudado por la recientemente desarrollada teoría de las cascadas electrón-fotón.

En 1948, Robert W. Williams, del Instituto de Tecnología de Massachusetts, (MIT) realizó un importante avance en el estudio de los chubascos aéreos. En lugar de tubos Geiger usó cámaras de ionización cuyas señales registró fotografiando las trazas en el osciloscopio. Trabajando en el Monte Evans en Colorado, colocó sus detectores en el centro y vértices de un triángulo equilátero de 12 metros de lado. Cuando un chubasco aéreo suficientemente grande incidía en su sistema de detectores o próximo a él, las partículas rá-



MAPA DE TIEMPO DE LLEGADA del mismo gran chubasco, registrado igualmente por algunos de los detectores del sistema de Rancho Volcano. (A fin de reducir costos, no todos los detectores estaban equipados para medir tiempo de llegada además de intensidad.) Las líneas diagonales en color indican los lugares donde el plano del disco cortó al plano del suelo en instantes sucesivos. Como el eje del chubasco era casi vertical, el plano del chubasco era casi horizontal y su línea de intersección con el suelo barrió el sistema de detectores en menos de un microsegundo (millonésima de segundo). La orientación del eje del chubasco se calculó a partir de los tiempos de llegada medidos haciendo pasar por ellos un plano de mejor adaptación. La concordancia entre los tiempos teóricos de llegada (*números de abajo*) y los medios (*números de arriba*) indica que la dirección del chubasco se calculó con un error de pocos grados. Los tiempos, en microsegundos.

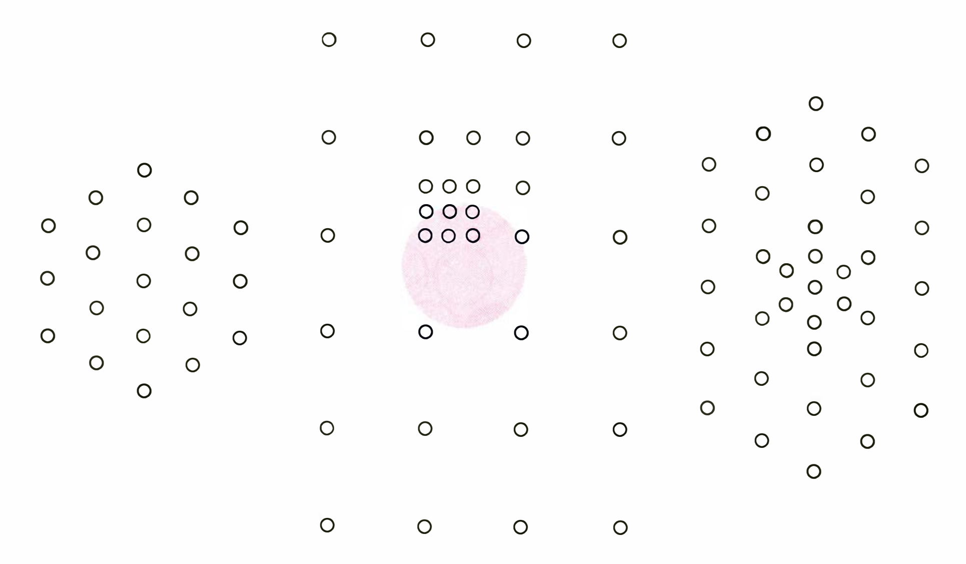
pidas producían señales en todas las cámaras de ionización prácticamente al mismo tiempo, de modo que el suceso se podía registrar de la manera usual. Las trazas osciloscópicas fotografiadas por Williams le revelaron algo más, sin embargo, que el hecho de que un chubasco había incidido en su aparato. Sabía que la amplitud de la señal procedente de cada cámara de ionización era proporcional al número de partículas que habían pasado a través de ella. Con esa información, Williams pudo localizar el eje del chubasco y calcular el número de partículas en él.

Cinco años más tarde, en 1953, la metodología de la observación de chubascos aéreos fue perfeccionada por Pietro Bassi, del MIT, trabajando con George W. Clark y Bruno Rossi. En lugar de cámaras de ionización usaron contadores de centelleo que estaban separados 30 metros y conectados a un sistema registrador electrónico de respuesta muy rápida. Fueron ellos, trabajando con este aparato, quienes demostraron que las partículas en la vecindad del eje de un chubasco se mueven formando un delgado disco plano. También llegaron a calcular las direcciones de movimiento de chubascos individuales a partir de las diferencias de tiempo de llegada que observaban entre pares de detectores montados sobre dos líneas que formaban ángulo recto.

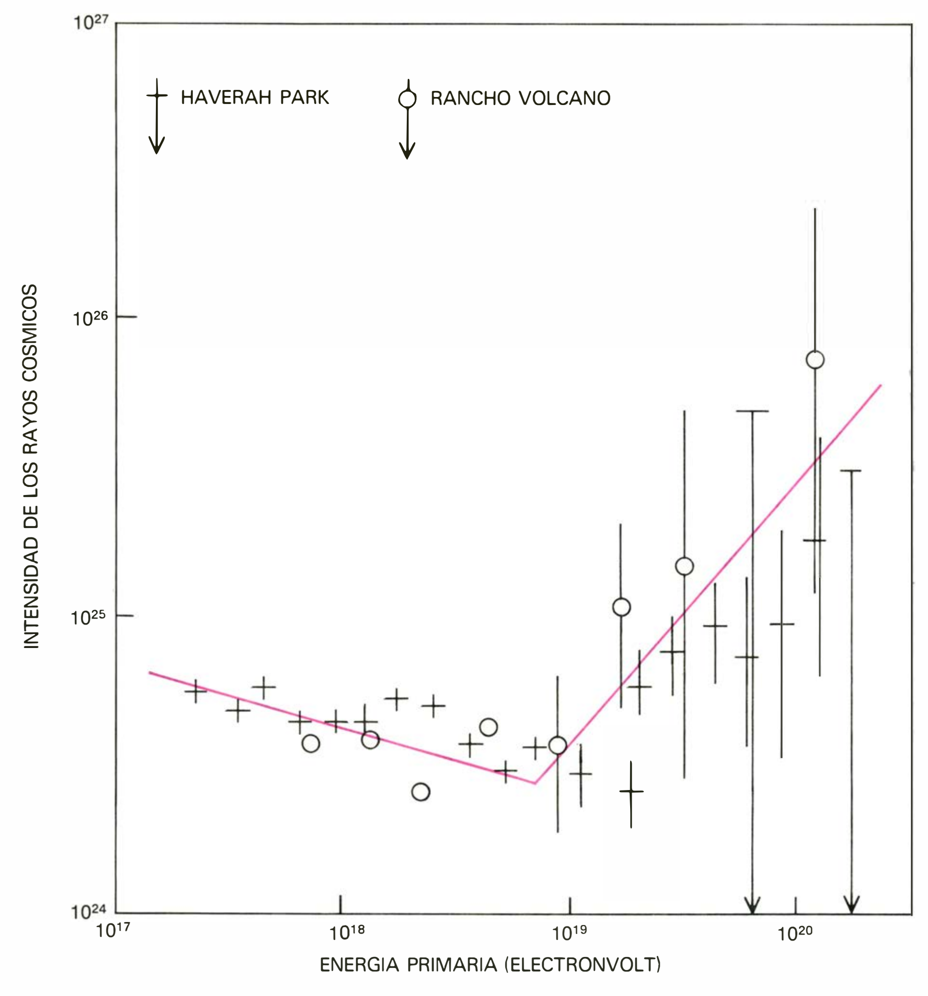
Los contadores de centelleo usados por Bassi y sus colaboradores se parecen a las cámaras de ionización en que la amplitud de la señal de salida es proporcional al número de partículas rápidas que la producen. Esta propiedad no se usó en el experimento que acabo de describir, pero no se olvidó la posibilidad de usarla en otro ulterior. Casi inmediatamente, se formó en el MIT un grupo más numeroso bajo la dirección de Rossi. Su objetivo era combinar la nueva técnica de medida de dirección con el método usado anteriormente por Williams para determinar el número de partículas en un chubasco. Hacia 1956, el grupo de Rossi estaba midiendo chubascos aéreos de rayos cósmicos con un sistema de 15 contadores de centelleo mucho más grandes en los alrededores de Boston, en la estación Agassiz del Observatorio del Harvard College. El área del sistema era de alrededor de una sexta parte de kilómetro cuadrado. Podía medir el número de partículas en chubascos con un error del 20 por cien y podía medir direcciones de chubascos con error de unos pocos grados. La técnica Williams-Bassi fue prontamente adoptada por casi todos los otros grupos que estudiaban chubascos aéreos.

El resultado más llamativo del sistema de Agassiz fue el descubrimiento, suministrado por un chubasco aéreo desusadamente grande, de que la energía de los rayos cósmicos primarios puede ser de al menos 5 x 10^{18} electronvolt. Este descubrimiento produjo cierta agitación en la comunidad astrofísica a causa de su relación con el problema del origen de los rayos cósmicos. Unos pocos años antes, la prueba de la existencia de los chubascos aéreos había sido decisiva para rechazar una idea propuesta en 1949 por R.D. Richtmeyer y Edward Teller: que todos los rayos cósmicos podían proceder del Sol. Richtmeyer y Teller argüían que la isotropía de los rayos cósmicos que se observaba (es decir, el hecho de que su intensidad fuera de la influencia de la Tierra parecía ser la misma en todas las direcciones) podía explicarse por la acción de un campo magnético hipotético que los confina (curvando sus trayectorias) dentro de un gran volumen que contiene al Sol. Concedían que no sería posible confinar partículas de energías mucho mayores que 10^{16} electronvolt sin suponer un volumen demasiado grande o un campo magnético demasiado intenso. Casi inmediatamente, otros investigadores hicieron observar que de los chubascos aéreos se deducía la existencia de muchos rayos cósmicos primarios con energías superiores a 10^{16} electronvolt, de modo que el Sol no podía ser la fuente de todos los rayos cósmicos.

Aunque ya no se acepta la hipótesis del origen solar de los rayos cósmicos, se continúan aduciendo argumentaciones semejantes a la usada por Richtmeyer y Teller cuando se discuten alternativas a la hipótesis del origen solar. Por ejemplo, la idea de que los rayos cósmicos están confinados por la acción de un campo magnético se puede aplicar a la totalidad de la galaxia. Según la hipótesis del origen galáctico, hay una energía crítica para los rayos cósmicos análoga a la energía crítica de 10^{16} electronvolt predicha por la teoría de Richtmeyer y Teller. El campo magnético de la galaxia parece ser algo más débil que el hipotético de Richtmeyer y Teller, pero el volumen de la galaxia es mucho mayor que el que ellos proponían. De aquí que la energía crítica de los rayos cósmi-



LOS SISTEMAS GIGANTES, tres de los cuales se esquematizan aquí, se designan usualmente por su localización. No existe una manera única y perfecta de disponer los detectores para formar un sistema. De izquierda a derecha, los diagramas de los sistemas en Rancho Volcano, en Sydney (Australia), y en Yakutsk, en la Unión Soviética. El círculo en color, que representa un chubasco aéreo con un diámetro efectivo de unos dos kilómetros, sirve para comparar la escala de los sistemas. El sistema de Rancho Volcano ya no existe en la forma representada. Tras cuatro años de funcionamiento (1959 a 1963) fue desmontado y reinstalado para formar una distribución más pequeña, y más densa, de las dos páginas precedentes.



ESPECTRO DE ENERGIA de los rayos cósmicos primarios más energéticos; se ha determinado mediante dos de los sistemas gigantes. El eje vertical representa la intensidad de los rayos cósmicos en términos de la función de Hillas (intensidad diferencial multiplicada por el cubo de la energía). Si el espectro tuviera la forma de una función potencial de exponente 1/3, los puntos experimentales estarían sobre una recta horizontal. En realidad, el espectro parece algo más pendiente que el correspondiente a la potencia 1/3 hasta una energía de 10^{19} electronvolt y más llano por encima de esa energía. La porción del espectro de energía donde la forma cambia se llama "el codo". (Los dibujos de las ilustraciones son obra de Ilil Arbel.)

cos para confinamiento galáctico es varios órdenes de magnitud superior a 10^{16} electronvolt.

La nueva plusmarca de energía de rayos cósmicos establecida por el mayor chubasco aéreo detectado por el sistema de Agassiz no excedía el "límite galáctico", mas para los defensores de la teoría del origen galáctico se le acercaba incómodamente. Casi tan pronto como entró en funcionamiento el sistema de Agassiz, comenzó en la Universidad de Cornell el trabajo de instalación de un sistema similar, pero cuya área de detección sería cinco veces mayor. En el MIT se hicieron planes para trasladar a Bolivia los detectores de Agassiz formando un sistema con un área cuatro veces superior a la original. Los resultados que se estaban obteniendo del experimento de Agassiz, sin embargo, confirmaban que la intensidad de la radiación cósmica decrece muy rápidamente con energía creciente. Unos sistemas que fueran sólo cuatro o cinco veces mayores que el de Agassiz no resolverían la cuestión de si los rayos cósmicos son tan energéticos que no pueden ser mantenidos dentro de la galaxia.

En calidad de nuevo miembro del grupo Rossi, yo me había estado entreteniendo en privado haciendo grandiosos planes para el más grande sistema que me parecía factible construir y poner en funcionamiento. Sería un sistema gigante, no cinco veces mayor que el de la Estación de Agassiz, sino cincuenta veces mayor. Yo no era el único con tales sueños en nuestro grupo, pero sí el único que en aquel momento tenía libertad suficiente para llevar adelante el plan. Tenía el apoyo de Rossi, que se preocupó de que se me suministraran todos los recursos materiales necesarios. Sobre todo, por medio de Rossi adquirí un inapreciable colega, Livio Scarsi, que estaba visitando el MIT con permiso de la Universidad de Milán. Hacia 1959, Scarsi y yo habíamos montado nuestro sistema gigante en Rancho Volcano, cerca de Albuquerque, Nuevo México.

Se había roto el hielo: se vio que los sistemas gigantes constituían una posibilidad real. Unos años más tarde, un grupo dirigido por G. Wilson, de la Universidad de Leeds, puso en funcionamiento un segundo sistema gigante en Haverah Park, cerca de Leeds. Desde entonces se han montado otros dos sistemas gigantes, uno en Australia por un grupo de la Universidad de Sydney encabezado por C.B.A. McCusker y el otro en Siberia cerca de la ciudad de Yakutsk, sólo unos grados al Sur del Círculo Polar Artico, por un grupo del Instituto de Investigación Cosmofísica y Aeronomía bajo la dirección de Yu. G. Shafer.

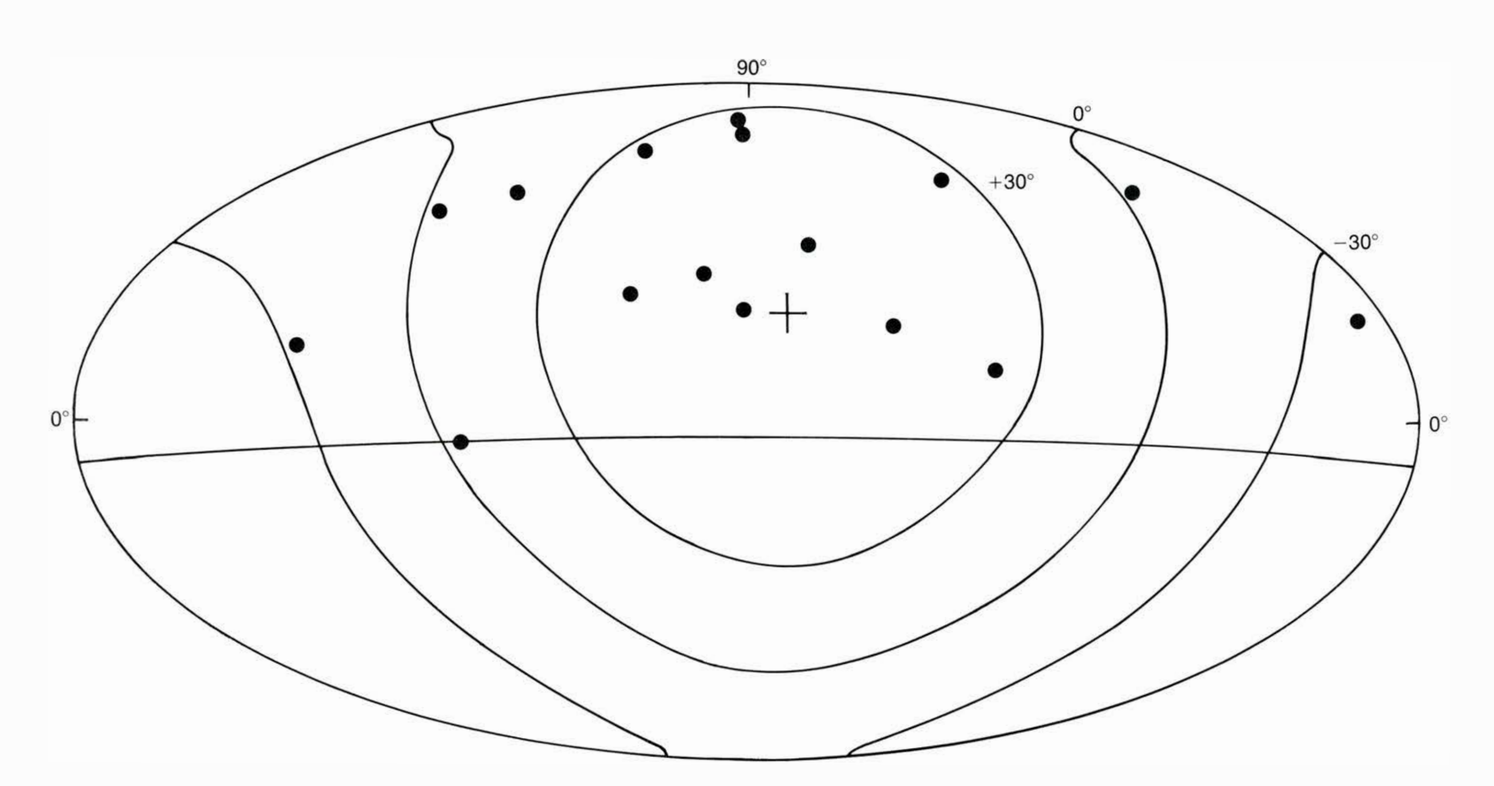
La idea de que los rayos cósmicos están confinados dentro de las galaxias había sido puesta en peligro en 1957 por el chubasco aéreo más energético observado con el sistema de Agassiz. En 1962 fue demolida por el mayor chubasco observado por Scarsi y yo en Rancho Volcano, chubasco cuya energía se estimó ser 10^{20} electronvolt. Cuando Scarsi y yo iniciamos nuestro proyecto no esperábamos detectar un chubasco de tan alta energía. Igualmente inesperado fue el descubrimiento de que había un número de partículas superior al predicho a todas las energías por encima de 10^{18} electronvolt. A esas altas energías, el espectro de rayos cósmicos se aplanaba, es decir, tenía un "codo". (El espectro de energía de los rayos cósmicos es una curva que describe la manera cómo la intensidad de los rayos cósmicos varía al ir hacia arriba o hacia abajo en la escala de energía.)

El achatamiento del espectro indicado por los datos de Volcano Ranch era un resultado que nadie había esperado. De hecho, los defensores de la teoría del origen galáctico había predicho justamente lo contrario. Preveían que el espectro de energía de los rayos cósmicos se haría más abrupto hacia las energías próximas a la crítica para confinamiento galáctico. ¿Sería posible que, por encima de cierta energía, la pérdida de rayos cósmicos por parte de nuestra galaxia sea más que compensada por rayos cósmicos que penetran en ella procedentes del espacio interplanetario? Si así fuese, ¿cuál es la fuente de los rayos cósmicos que penetran en ella? ¿Son los quasars o cualesquiera otros objetos conocidos en astronomía, capaces de rellenar las vastas extensiones de espacio entre galaxias con partículas subatómicas tan notablemente energéticas?

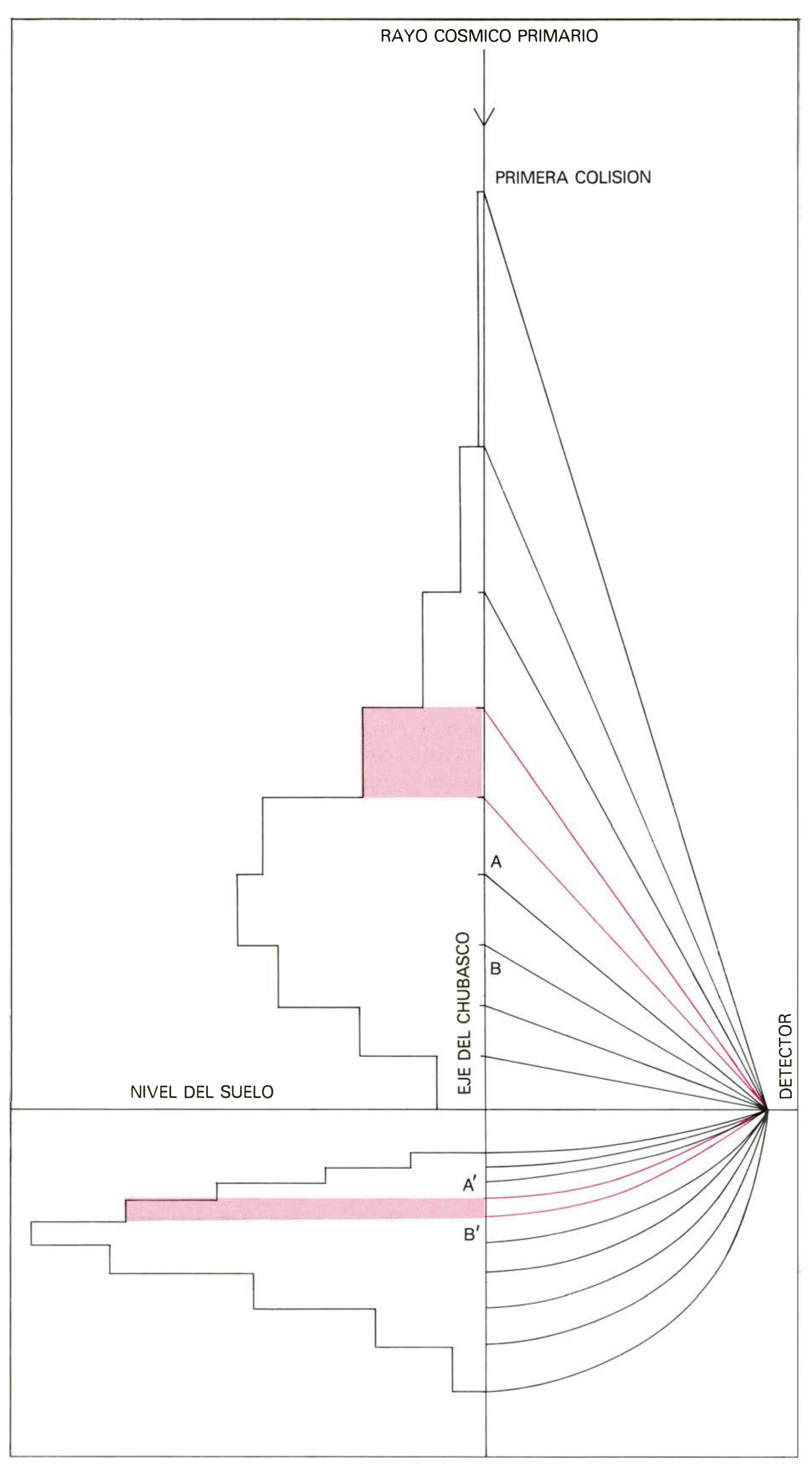
El suceso de 20^{20} electronvolt observado en Rancho Volcano adquirió pronto una importancia adicional. En 1965, Arno A. Penzias y Robert W. Wilson, de los Laboratorios Bell, descubrieron una radiación electromagnética peculiar, en la banda de frecuencia de las microondas, que al parecer llena la totalidad del universo. Algunas veces se le llama "radiación de desecho" porque se cree ser la radiación que quedó como resultado de la "gran explosión" con la que comenzó el universo. Al enterarse del descubrimiento de Penzias y Wilson, Kenneth Greisen de la Universidad de Cornell y G.T. Zatsepin del Instituto de Física P.N. Lebedev de Moscú, se dieron cuenta independientemente de que el encuentro entre los rayos cósmicos y los fotones de la radiación de desecho reduciría considerablemente la energía de los rayos cósmicos por encima de cierto nivel umbral de energía que pudieron calcular fácilmente. Mostraron que si las fuentes de rayos cósmicos están uniformemente distribuidas en el universo, la consecuencia de esos encuentros sería un brusco escarpamiento del espectro de energía de los rayos cósmicos hacia los 6×10^{19} electronvolt. La intensidad de los rayos cósmicos de energía igual o superior a 10^{20} electronvolt se reduciría tanto que los chubascos de 10^{20} electronvolt deberían ser imposibles de detectar.

¿Podríamos Scarsi y yo habernos equivocado acerca de la energía de nuestro chubasco más grande? Parece que no. A medida que los otros sistemas gigantes empezaron a funcionar, también en ellos se registraron chubascos de energías tan altas como el nuestro e incluso más altas. El codo en el espectro también se ha confirmado. ¿Qué panorama sugieren esos resultados sobre el origen de los rayos cósmicos y qué otras clases de información pueden suministrar los sistemas gigantes?

Se cree en general que prácticamente todos los rayos cósmicos en el espectro que están por debajo del codo se origina en fuentes dentro de nuestra galaxia. Además, es atractivo suponer que esos rayos cósmicos están confinados en la galaxia, porque en ese caso las hipotéticas fuentes de los rayos cósmicos no necesitan ser tan potentes como habrían de serlo en caso contrario. Si los rayos cósmicos están



DIRECCIONES DE LLEGADA de los 16 chubascos más energéticos registrados por el sistema de Haverah Park desde 1963, que suministran la más convincente prueba de que los rayos cósmicos de máxima energía se originan fuera de nuestra galaxia. Cada chubasco tenía una energía superior a 5×10^{19} electronvolt. Las direcciones individuales aparecen representadas en una proyección de Aitoff de la esfera celeste, que conserva las áreas. La porción del cielo que está por encima de la línea que atraviesa la parte central de la proyección pasa sobre el sistema y es explorada por éste una vez cada 24 horas sidéreas. La cruz por encima del centro representa el polo norte galáctico, que está inclinado unos 60 grados respecto al eje de la Tierra. También se indican los paralelos correspondientes a las latitudes galácticas de −30 grados, 0 grados y + 30 grados. En la hipótesis de que los 16 rayos cósmicos de máxima energía se hubieran originado en la galaxia, muchos menos de ellos deberían haber llegado procedentes de direcciones dentro de 30 grados del polo norte galáctico.



DESARROLLO LONGITUDINAL de un chubasco aéreo, reflejado en la forma del impulso producido en un detector situado a un lado del eje del chubasco. El perfil longitudinal de un chubasco se puede medir detectando la radiación emitida por el chubasco al pasar a través de la atmósfera; tal perfil es una rica fuente de información acerca de la partícula que dio origen al chubasco. El principio se ilustra bajo tres hipótesis simplificadoras: primero, que la región desde la que se emiten las partículas detectadas tiene un tamaño despreciable; segundo, que se mueve a lo largo del eje del chubasco con la velocidad de la luz, y tercero, que la radiación detectada se mueve en líneas rectas también a la velocidad de la luz. El histograma por encima del nivel del suelo es el perfil de emisión, representando las contribuciones relativas a la señal luminosa de sucesivos intervalos a lo largo del eje del chubasco. El histograma de abajo es el correspondiente perfil del tiempo de llegada. Como todas las partículas se mueven en la misma velocidad, los intervalos de tiempo son proporcionales a las distancias. Se usa en los detectores de muones y en los detectores de luz para recibir la luz Cerenkov atmosférica (haz dirigido hacia adelante producido por efecto Cerenkov).

confinados, su intensidad en el espacio interestelar va aumentando hasta que se equilibra con la pérdida a través del espacio intergaláctico (las pérdidas en el interior de la galaxia son poco importantes). Como el supuesto confinamiento se debe a campos magnéticos, no es igualmente efectivo para rayos cósmicos con distintas energías. Los rayos cósmicos de baja energía son confinados muy efectivamente mientras que aquellos cuyas energías se acercan a la crítica tienen una creciente tendencia a escapar de la galaxia y aquellos de energía superior a la crítica ya no quedan en absoluto confinados.

No hay contradicción entre la idea del confinamiento y la teoría de los dos componentes sobre el origen de los rayos cósmicos, según la cual los rayos cósmicos por debajo de cierta energía son de origen predominantemente galáctico mientras que los que la tienen superior son predominantemente extragalácticos. Muchos estudiosos de la teoría de los rayos cósmicos se sintieron atraídos hacia el modelo de dos componentes cuando los experimentos de chubascos aéreos mostraron que el espectro de los rayos cósmicos, que se suponían cada vez más abrupto al acercarse a la energía crítica, se hacía por el contrario más llano. Otros teóricos, sin embargo, continuaron defendiendo la hipótesis de que todos los rayos cósmicos se generan dentro de la galaxia. Indicaron que hay más de una manera de compensar la pérdida de los rayos cósmicos de alta energía hacia el espacio intergaláctico. Se sabe poco acerca de los aceleradores naturales que generan la radiación cósmica: quizás el espectro de rayos cósmicos se achata porque el rendimiento de los aceleradores aumenta dramáticamente por encima de cierta energía, más que compensando cualesquiera aumentos de pérdidas hacia fuera de la galaxia. Así, la forma del espectro de energía de los rayos cósmicos no es crucial para decidir entre la teoría del origen galáctico y la de los componentes: debe buscarse alguna otra clase de indicación para elegir entre una y otra. Las medidas de dirección de los chubascos aéreos suministran la clase de indicación que se necesitaba.

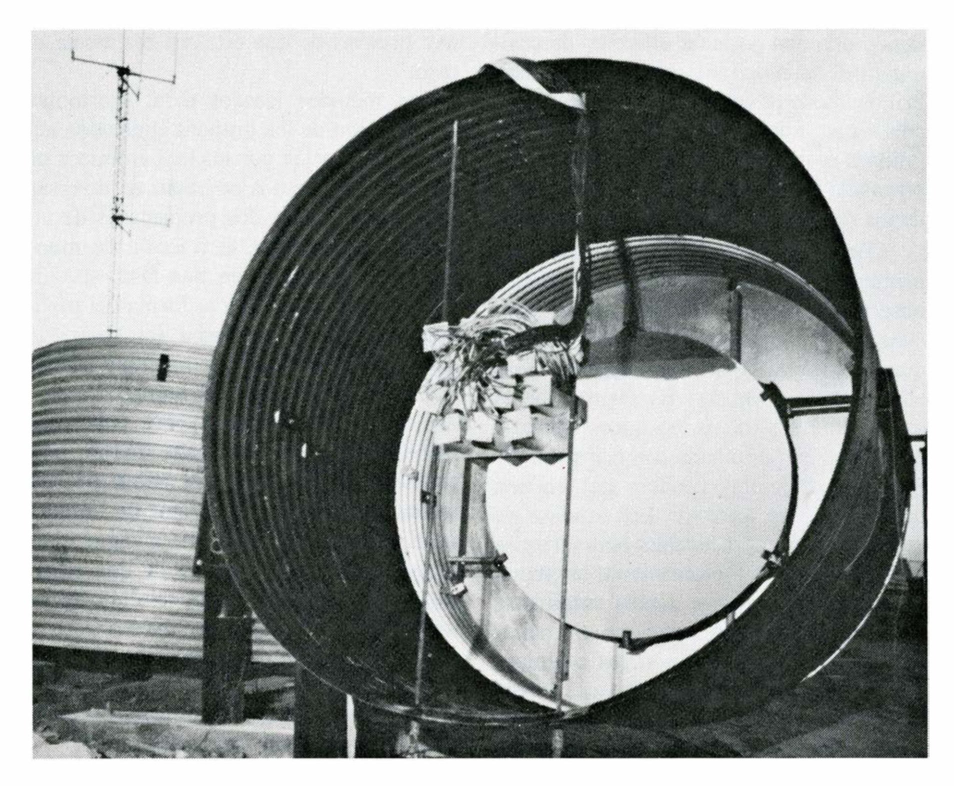
La teoría del confinamiento explica el que los rayos cósmicos de energías bajas y moderadamente altas sean casi isótropos: vienen de todas las direcciones del espacio con intensidades casi iguales. Por encima de la energía crítica, sin embargo, los rayos cósmicos no están en absoluto confinados y no pueden siquiera ser muy desviados por el campo magnético galáctico; por tanto, sus direcciones de llegada deben indicar hacia dónde están situadas sus fuentes. Si incluso los rayos cósmicos

de máxima energía se originan en la galaxia, entonces la mayor parte de ellos deben llegar procedentes del plano de la Vía Láctea y muy pocos de direcciones más o menos perpendiculares a tal plano. Si se originan en galaxias activas fuera de la nuestra, se puede localizar, con el tiempo, las más importantes galaxias emisoras de rayos cósmicos.

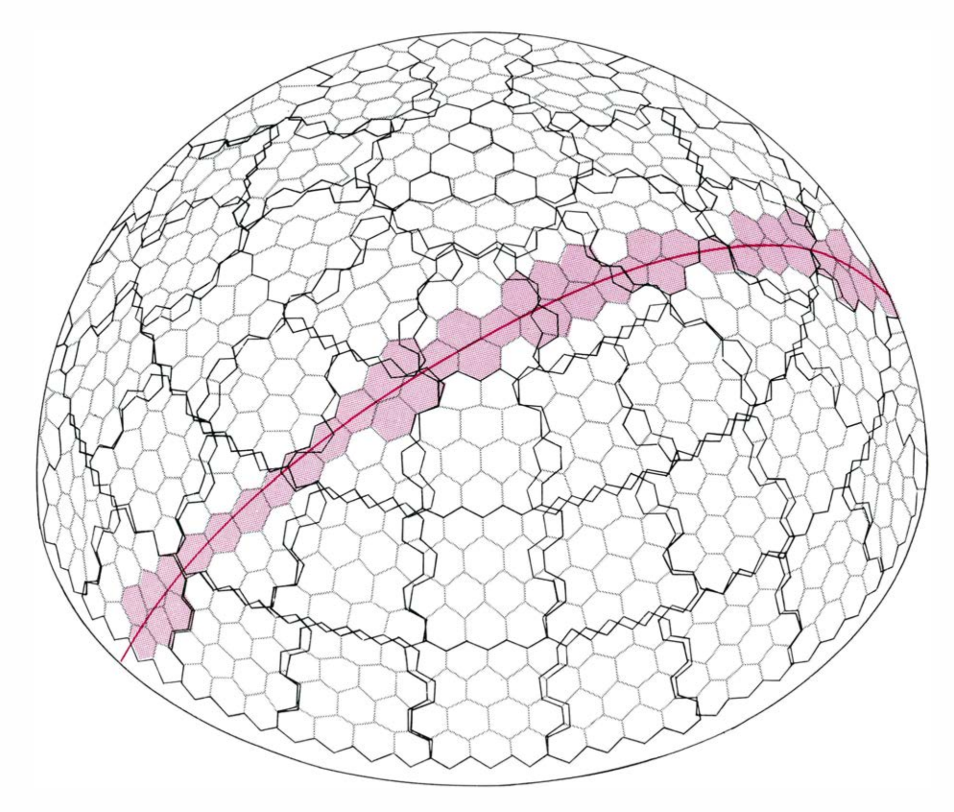
Es interesante pensar sobre lo que puede quedar revelado cuando se hayan acumulado más datos particularmente a energías por encima de 6 x 10^{19} electronvolt. Parece virtualmente cierto que el espectro de energía debe mostrar alguna clase de estructura que se puede relacionar con el efecto Greisen-Zatsepin, suministrando así una comprobación acerca de la exactitud de las actuales medidas de energía en este dominio, las cuales se basan en el análisis detallado de la estructura de los chubascos aéreos. El efecto Greisen-Zatsepin tiene también la posibilidad de suministrar una escala independiente de distancia astronómica cubriendo el dominio de 30 millones a 1000 millones de años-luz.

Para poder ser tan optimista, sin embargo, he tenido que suponer que va a resolverse un problema importante, un problema al que hasta ahora no he dedicado la atención que merece. Mis comentarios sobre la interpretación de los datos existentes, particularmente los datos sobre la dirección de llegada de los rayos cósmicos, se basan en una fórmula bien conocida que permite calcular la curvatura de la trayectoria de una partícula cargada que se mueve perpendicularmente a un campo magnético. Si la partícula es altamente relativista, esto es, si se mueve con una velocidad muy próxima a la de la luz, el radio de curvatura es proporcional al cociente de su energía total a la intensidad del campo. En el caso de rayos cósmicos, las medidas de los chubascos aéreos dan la energía total de la partícula y los datos astronómicos proporcionan una estima de la intensidad del campo magnético. Pero el radio de curvatura es también inversamente proporcional a la carga eléctrica de la partícula en movimiento, que para rayos cósmicos de baja energía está probablemente comprendido entre 1 (para protones, o núcleos de hidrógeno) y 26 (para núcleos de hierro).

Así, si la energía crítica para confinamiento galáctico de protones de rayos cósmicos es, de acuerdo con cierto modelo, de 2,6 x 10^{19} electronvolt, la energía crítica para confinamiento de núcleos de hierro de rayos cósmicos según el mismo modelo será 26 veces menor, o sea solamente 10^{18} electronvolt. Por tanto, son cosas muy diferentes si, por una parte, los rayos cósmicos de máxima energía son protones o núcleos de hierro o, por otra,



UN ELEMENTO del nuevo detector de chubascos "Ojo de Mosca", en construcción en la Pequeña Montaña de Granito, en las llanuras saladas de Utah. Consta de un espejo de 1,5 metros de diámetro. El detector completo estará compuesto de 67 de tales espejos, cada uno con 14 o 16 tubos fotomultiplicadores montados en su plano focal. (El espejo que aparece aquí tiene 14.) Cada tubo fotomultiplicador detectará y registrará débil luz emitida por los chubascos de rayos cósmicos en la atmósfera. Cuando estén completados los 67 espejos, se dispondrán de modo que formen un "ojo" compuesto, de 880 tubos fotomultiplicadores.



EL DETECTOR OJO DE MOSCA, cuando esté terminado, vigilará el cielo en busca de chubascos de rayos cósmicos. Cada uno de los polígonos ligeramente solapados, delimitado en negro, representa el campo de visión de uno de los 67 espejos. Cada polígono está formado bien por 12 o por 14 hexágonos, cada uno de los cuales representa el campo de visión de seis grados de un tubo fotomultiplicador. Un chubasco que pase sobre el Ojo de Mosca emitirá una débil luz que activará el fotomultiplicador a través de cuyo campo de visión pase (*color*). Cada tubo fotomultiplicador tendrá su propio canal electrónico para registrar tiempo e intensidad de las señales recibidas. A partir de éstas, un ordenador reconstruirá el perfil del chubasco.

tienen una composición diferente de cualquiera de esas dos alternativas. El descubrimiento de la composición de las partículas que inician los chubascos aéreos grandes es un problema no resuelto de importancia fundamental en la investigación de los rayos cósmicos de máxima energía.

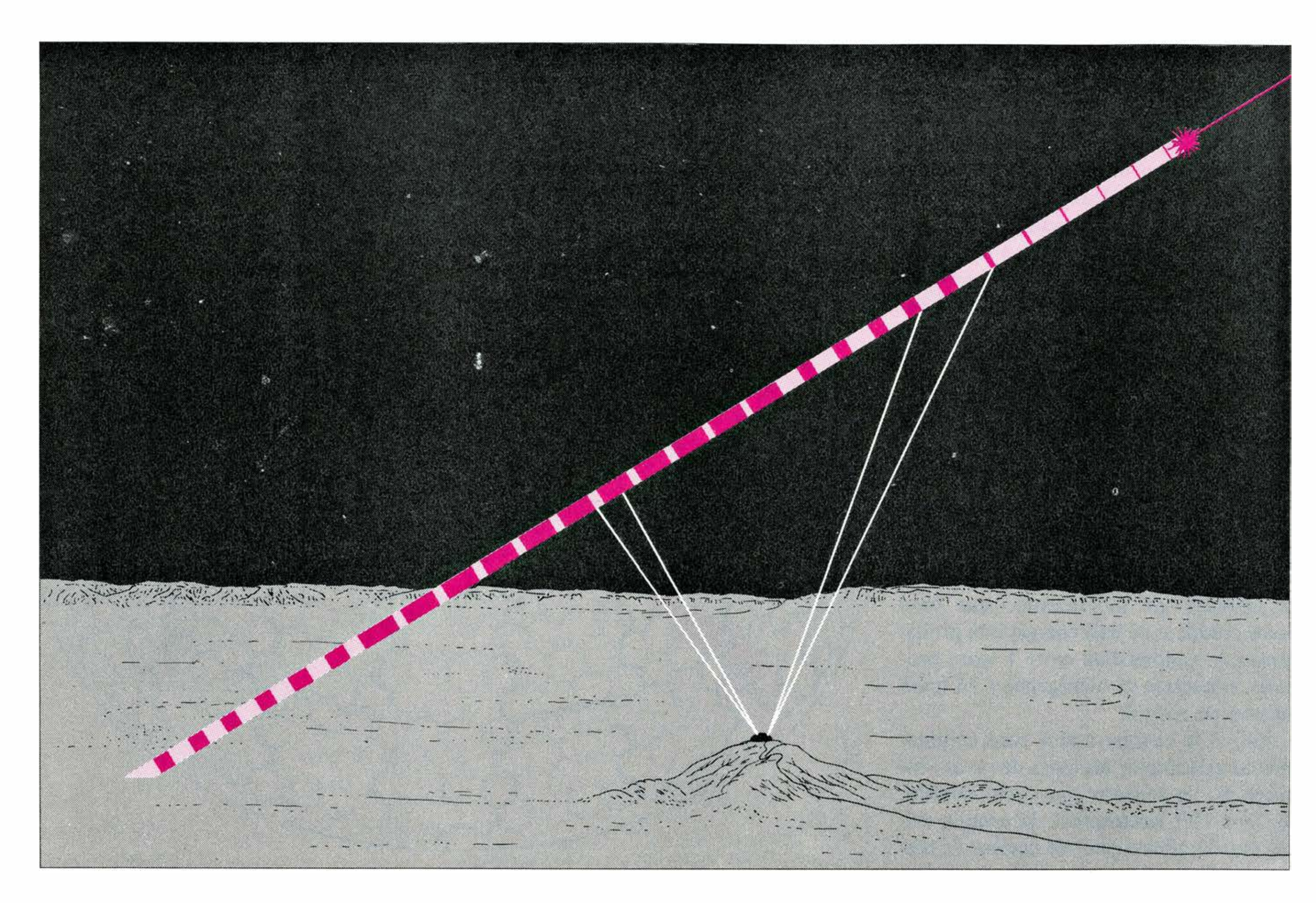
¿Cómo se puede saber qué partículas primarias producen los chubascos aéreos observados? No hay duda de que las partículas de baja energía que dan lugar a pequeños chubascos son virtualmente todas ellas núcleos atómicos. Es seguro que es así porque los rayos cósmicos de baja energía se han estudiado con cierto detalle mediante detectores situados por encima de la atmósfera terrestre. Los estudios de esta clase son impracticables para energías por encima de 10^{15} electronvolt; la estructura de los chubascos aéreos suministra pruebas, sin embargo, de que al menos una gran mayoría de los rayos cósmicos que dan lugar a grandes chubascos aéreos son también núcleos atómicos. Aunque es posible que algunos de los rayos cósmicos de máxima energía sean partículas subatómicas de menor masa que los núcleos, no hay pruebas de que éste sea realmente el caso.

Los métodos ideados para determinar qué fracción de los grandes chubascos aéreos es producida por núcleos atómicos de diferentes elementos se basan en diversas maneras de medir dos propiedades de un chubasco individual: la fracción de muones en un chubasco en una fase especificada de su desarrollo y la forma del perfil longitudinal del chubasco (es decir, la curva que describe el crecimiento y la subsiguiente disminución del número de partículas en el chubasco). Hay poderosas razones para creer que cada una de esas propiedades variará de una forma característica, dependiendo del factor de Lorentz de la partícula primaria que inició el chubasco: el factor de Lorentz de una partícula en movimiento es la relación de su energía total a su masa en reposo. Cuanto mayor sea el factor de Lorentz de la partícula primaria, tanto menor será la fracción de muones en el chubasco y tanto más ancho el perfil del chubasco.

La fracción de muones en un chubasco se puede medir mediante un sistema de detectores del tipo Williams-Bassi, que mide el número de partículas de todas clases, junto con un detector de partículas rápidas protegido por un espesor suficiente de cemento, plomo o tierra que absorba todos los electrones y fotones, dejando solamente los muones. Los métodos para registrar directamente perfiles de chubascos individuales constituyen un avance reciente sobre el cual volveré más adelante. En el pasado solamente ha sido posible buscar diferencias entre alguna propiedad que esté correlacionada con el perfil.

Si se examinan esas propiedades para todos los tamaños de chubascos que se han observado, puede decirse que, a medida que la energía total del chubasco aumenta, el factor de Lorentz lo hace aproximadamente al mismo paso. Por tanto se puede asegurar que no existe cambio radical en la masa de las partículas primarias que originan chubascos de distintos tamaños.

La teoría que relaciona las dos propiedades con el factor de Lorentz de la partí-



LA TRAYECTORIA de un chubasco aéreo se puede calcular mediante datos que se registraran con el Ojo de Mosca. Aquí se dibuja el detector ya terminado. Como la luz de un chubasco aéreo es tan débil, solamente podrá detectarse por la noche. Las bolas de fuego de los chubascos aéreos se mueven en línea recta a virtualmente la velocidad de la luz. Las cuatro líneas que van desde la trayectoria del chubasco (*en color*) al detector representan los caminos por los cuales la luz procedente del chubasco llegará a dos pares separados de tubos fotomultiplicadores adyacentes en el detector. La diferencia de tiempo entre las señales de fotomultiplicadores contiguos variará de una manera característica: depende de la relación geométrica entre el detector y la trayectoria.

cula primaria no es exacta, sin embargo, ni lo son nuestras medidas. Con los métodos actuales es imposible distinguir entre núcleos diferentes con la misma energía, ni siquiera entre un protón y un núcleo de hierro. El problema es realmente serio. Aun cuando se conozca la energía de un rayo cósmico y la dirección de donde procedía cuando entró en la atmósfera, ¿cómo se puede reconstruir su trayectoria magnéticamente curvada en el espacio interestelar, si su carga no se conoce con precisión superior a un factor de 26? No debe olvidarse que por encima de una energía de 10^{15} electronvolt, los rayos cósmicos pueden continuar suministrando indefinidamente el único medio que tienen los investigadores para estudiar interacciones nucleares. ¿Cómo puede uno interpretar lo que sucede en colisiones entre partículas a esas energías sin saber qué partículas constituyen el haz incidente?

Hay en la actualidad varios nuevos experimentos en curso con el fin de medir directamente el perfil de los chubascos a medida que se desarrollan en la atmósfera. Recordemos el modelo del pequeño disco deslizando enhebrado en un hilo. Las partículas dentro del disco producen luz. El disco del chubasco aéreo es por tanto una especie de bola de fuego cuya luminosidad aumenta a medida que crece el número de partículas, alcanza un brillo máximo al llegar el chubasco a su plenitud y empieza después a apagarse. Los más modernos métodos de medida de perfiles de chubascos aéreos son conceptualmente equivalentes a fotografiar la bola de fuego móvil por la noche mediante la propia luz emitida por ella. El registro del chubasco aéreo suministrado por esos métodos equivale a una sucesión de fotografías instantáneas.

La luz de la bola de fuego consta de dos componentes correspondientes a dos procesos diferentes mediante los cuales las partículas rápidas pueden producir luz. Una componente, que en este caso es con mucho la más intensa, es proyectada hacia adelante como la luz del faro frontal de una locomotora. La componente menos intensa se emite con la misma intensidad en todas las direcciones, como la luz procedente de un meteoro. El haz dirigido hacia adelante se produce por efecto Cerenkov y se llama luz Cerenkov atmosférica. La luz de emisión isótropa se produce por fluorescencia. Algunos de los nuevos experimentos usan la luz Cerenkov atmosférica producida por los chubascos aéreos; otros usan la luz fluorescente.

La técnica para fotografiar chubascos aéreos mediante la luz Cerenkov atmosférica ya ha sido perfeccionada y está en uso en dos de los sistemas gigantes, el de Inglaterra y el de Siberia. El mejor ángulo para obtener exposiciones intensas es aquel en que el chubasco se aproxima a la cámara de frente. Sin embargo, la bola de fuego se mueve virtualmente con la misma velocidad de la luz que emite, de modo que el ángulo entre el eje del chubasco y la visual de la cámara no puede ser demasiado pequeño porque resultaría imposible resolver el perfil del chubasco. El mejor compromiso entre intensidad de la señal y poder de resolución es el dado por chubascos que hacen blanco a varios cientos de metros de la cámara. El colector de luz empleado en un experimento de este tipo tiene poca semejanza con una cámara tomavistas ordinaria. El colector carece de componentes ópticos. Las exposiciones sucesivas se superponen espacialmente; se separan en el dominio temporal.

Después que las trayectorias del chubasco se han calculado de la manera corriente a partir de los datos proporcionados por el sistema gigante, el perfil espacial aparente se puede deducir del perfil temporal mediante un sencillo método gráfico. Deducir el perfil verdadero, sin embargo, es mucho más difícil y requiere el uso de un modelo teórico detallado de las cascadas de nucleones, con la correspondiente incertidumbre en los resultados.

La técnica para fotografiar los chubascos aéreos mediante la luz fluorescente atmosférica está ahora en elaboración en los Estados Unidos y Japón. Una ventaja de este enfoque es que los chubascos aéreos se fotografían lateralmente, con un ángulo ideal para la buena resolución espacial. Otra ventaja es que mediante este método debe ser posible detectar chubascos aéreos a una distancia considerable, tal vez decenas de kilómetros. La cámara que se necesita para fotografiar la luz fluorescente debe ser mucho más complicada que la que se usa para registrar la luz Cerenkov; por otra parte, las trayectorias de los chubascos se pueden obtener a partir de un análisis de sus fotografías. De aquí que la cámara para fluorescencia atmosférica es un instrumento independiente, que no necesita información suplementaria de un sistema detector adyacente del tipo Williams-Bassi. Finalmente, y quizás ello sea lo más importante, sólo se necesitan pequeñas correcciones para deducir el perfil verdadero de un chubasco a partir de fotografías de su luz fluorescente.

Las grandes ventajas potenciales de la detección de chubascos aéreos mediante la fluorescencia atmosférica se conocían al menos desde 1962. También se conocían sus dificultades, la principal de las cuales es la extremadamente baja intensidad de la luz fluorescente. Para lograr siquiera un éxito limitado, se necesitarían numerosos colectores de luz grandes y de alto rendimiento en una situación casi ideal, con aire limpio, condiciones meteorológicas favorables y una luminosidad de fondo tan pequeña como fuera posible. Entre 1965 y 1974 se realizaron varios intentos fallidos para construir un sistema práctico. Hace dos años, un grupo de la Universidad de Utah dirigido por Haven E. Bergeson trajo tres nuevos colectores de luz altamente sensibles a Rancho Volcano y realizaron con éxito varias pruebas en conjunción con el sistema de Rancho Volcano. Está ahora en construcción un dispositivo completo consistente en 67 colectores luminosos cuyo campo de visión está dividido entre 880 tubos fotomultiplicadores. A medida que los colectores van siendo acabados de construir, se van instalando en la Pequeña Montaña de Granito, en los Campos Experimentales Dugway a unos 160 kilómetros al oeste de la Ciudad del Lago Salado.

El dispositivo de Utah ha sido bautizado Ojo de Mosca. Cada tubo fotomultiplicador vigila su propio y estrecho campo hexagonal de visión, de la misma manera que lo hace cada elemento en el ojo compuesto de una mosca. Para facilidad de operación, los colectores se están instalando separadamente, cada uno en un alojamiento giratorio propio. A la caída de la noche, con buen tiempo, los colectores se colocan automáticamente en posición, de manera que juntos observan la bóveda celeste completa. Las señales procedentes de los tubos fotomultiplicadores serán examinadas por un ordenador en línea con ellos. La "traza" de un chubasco aéreo es una bola de fuego moviéndose en línea recta a la velocidad de la luz. Cuando el ordenador reconoce un suceso con esa traza, dará a una cámara tomavistas de alta velocidad la orden de que registre el suceso, en clave, para ulterior análisis. Del registro así obtenido se puede reconstruir la trayectoria del chubasco en las tres dimensiones en el espacio. La reconstrucción dará inmediatamente la energía y la dirección de llegada del rayo cósmico primario que originó el chubasco.

Si el funcionamiento del Ojo de Mosca responde a lo que de él se espera, el dispositivo acumulará datos sobre los rayos cósmicos de máxima energía a una velocidad 10 veces superior a la de todos los sistemas gigantes existentes juntos. Los perfiles de chubascos medidos directamente que también proporcionará deben jugar un importante papel en la resolución del problema de la composición, superando así uno de los últimos obstáculos que quedan para descubrir el origen de los rayos cósmicos de máxima energía.

Ciencia y sociedad

Cinética celular

La introducción de trazadores radiactivos en los años 40 condujo a importantes avances en biología celular. Uno de los más significativos fue, sin duda, el análisis del ciclo de reproducción celular mediante la utilización de trazadores de la síntesis de ADN. En 1951, Howard y Pelc, utilizando fósforo-32 como marcador en raíces de *Vicia faba*, demostraron que el ADN es sintetizado en un intervalo discreto dentro de la interfase. Posteriormente, otros investigadores, empleando timidina tritiada como marcador en diferentees tipos celulares, llegaron a resultados parecidos. En breve, la síntesis del ADN cromosómico, que necesariamente precede a la división celular, no ocurre de forma continua en toda la interfase, sino únicamente en un intervalo definido de la misma, con lo cual, el ciclo celular queda dividido en cuatro intervalos. El intervalo de síntesis del ADN*(S)*, el intervalo *G*2 o lapso entre el final del intervalo *S* y la división celular, el intervalo de división o mitosis *(M)* y el intervalo *G*1, que queda interpuesto entre el final de la mitosis y el comienzo del intervalo *S*.

La duración de los intervalos *S*, *G*2 y *M* es relativamente invariante para un tipo celular en distintas condiciones ambientales e, incluso, para distintas clases de células. El intervalo *G*1, por el contrario, es extremadamente variable y, de hecho, su duración es la que determina la velocidad de proliferación celular. La extrema variabilidad del *G*1 plantea un problema fundamental en relación con la estructura cinética del ciclo celular. De acuerdo con la concepción clásica, el ciclo sería un programa metabólico invariante y circular, constituido por un encadenamiento determinista de sucesos, de suerte que, dada la posición de una célula en el ciclo, podrá conocerse el tiempo que tardará en llegar a cualquier otro punto del mismo. Si bien este esquema es aplicable a los intervalos *S*, *G*2 y *M*, no lo es al intervalo *G*1.

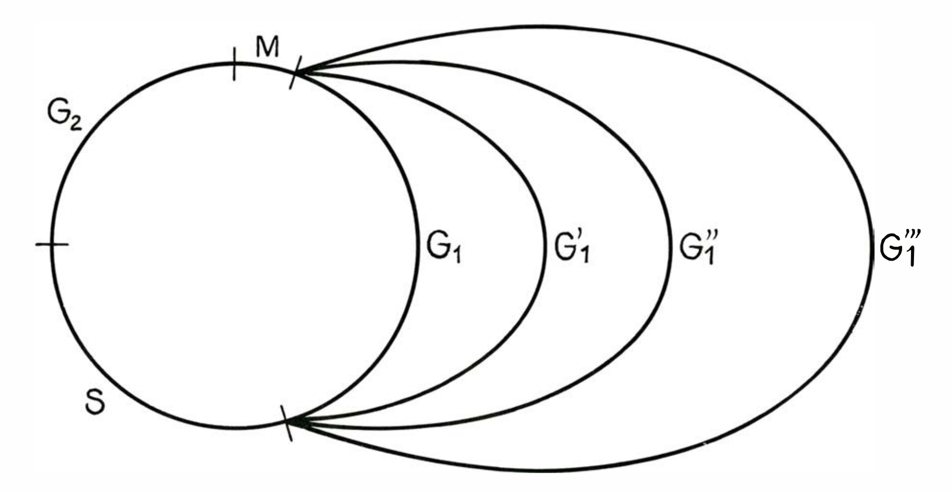
Mediante microcinematografía aplicada a cultivos celulares, varios autores pudieron medir la duración del ciclo en células individuales de una misma población y obtuvieron las correspondientes distribuciones de frecuencias. El resultado fundamental de estos estudios fue la prueba de una gran variabilidad de la duración del ciclo celular, imputable al *G*1, dentro de una misma población e, incluso, entre células hermanas (procedentes de la misma división) y que la distribución acumulativa de frecuencias era una exponencial característica del tipo celular y de las condiciones de cultivo. Esto fue interpretado por Smith y Martin, del Imperical Cancer Research Fund de Londres, mediante un nuevo modelo del ciclo celular que dieron a conocer en 1973 *(Proc. Nat. Acad. Sci. USA* 70, 1263-1267). Este modelo, denominado de probabilidad de transición, admite que las células, algún tiempo después de la mitosis, entran en un estado en el que sus actividades no están dirigidas a la reproducción (estado *A*), pudiendo permanecer un tiempo no definido en dicho estado, pero existiendo una probabilidad constante para abandonarlo entrando en el programa de reproducción (fase *B*) que comprendería una parte del *G*1 y los intervalos *S*, *G*2 y *M*. De esta manera, el ciclo celular queda dividido en dos partes: Una, completamente determinista y de duración constante (la fase *B*) y otra, de duración indeterminada (el estado *A*) que representa una discontinuidad en el programa de reproducción, el cual se reinicia al azar pero con una probabilidad de transición definida para el tipo celular y condiciones ambientales. Este modelo constituye una generalización, a todas las clases de células, de la hipótesis de un estado de reposo o *G*0, introducido por Lajtha en 1963 *(J. Cell. Comp. Physiol.* 62, 143-145) para explicar la cinética de reproducción en la regeneración hepática, perfeccionada y ampliada a todas las poblaciones de proliferación lenta por Burns y Tannock en 1970 *(Cell Tissue Kinet.* 3, 321-334). La gran variabilidad de la duración del ciclo entre células de la misma población, e incluso, entre hermanas, no solamente es compatible con el nuevo modelo, sino que es la consecuencia inevitable de la transición aleatoria que propone.

Con respecto a la economía celular, la hipótesis de Smith y Martin es mucho más satisfactoria que la concepción clásica del *G*1. En efecto, todas las células tendrían el mismo programa de reproducción, o muy similar, y las diferencias de velocidad de proliferación no serían debidas a cambios en el número de pasos del programa sino a la frecuencia con que éste se inicia. Se trata, por tanto, de una regulación del tipo *interruptor conectado-desconectado.*

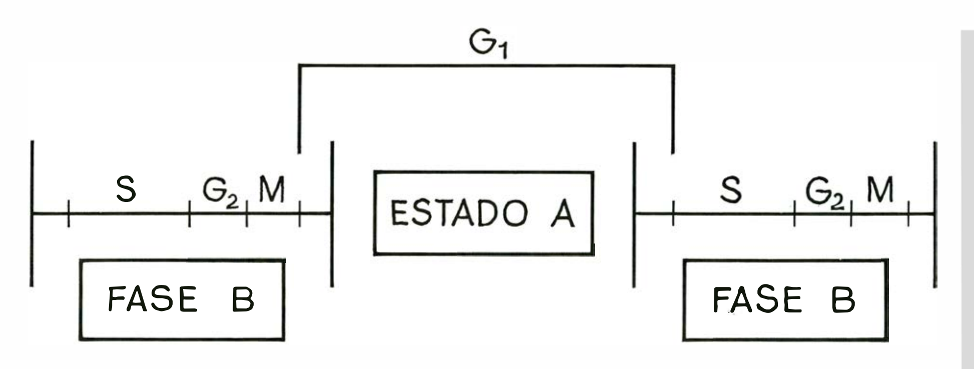
Respecto a la base molecular de la transición del estado *A* a la fase *B*, los autores del modelo propusieron la siguiente hipótesis: La transición dependería de que uno o más factores iniciadores de la síntesis del ADN cromosómico, cuya concentración oscilara irregularmente, alcanzara un valor crítico o umbral. La probabilidad de transición dependería de la diferencia entre el valor umbral y la concentración media del iniciador, así como de la amplitud de sus oscilaciones. Hay pruebas tanto teóricas como experimentales de que pueden darse fluctuaciones de este tipo para diferentes componentes de la célula.

A pesar de las pruebas experimentales favorables al modelo, existen algunos puntos débiles que han sido sacados a la luz por diferentes autores desde la publicación del modelo. Ello ha dado lugar a vivas controversias que han resultado muy estimulantes, provocando comprobaciones cada vez más rigurosas. Uno de los puntos débiles consistía en que, si bien el modelo no estaba en contradicción con datos cinéticos *in vivo*, tampoco existía una prueba directa de que fuera aplicable a esta situación. Realmente la única prueba directa eran los estudios sobre los tiempos intermitóticos de células individuales medidos mediante microcinematografía en cultivos celulares.

Cabía, por tanto, la posibilidad de que la transición aleatoria fuera un artefacto *in vitro*. Este problema ha sido abordado re-



El ciclo celular según la concepción clásica. Obsérvese la variabilidad del intervalo G1



El ciclo celular según el modelo de probabilidad de transición propuesto por Smith y Martin

cientemente por nuestro grupo en la Cátedra de Histología de la Facultad de Medicina (Universidad de Barcelona) mediante la medición en hígado de embrión de pollo de un nuevo parámetro cinético, ideado por nosotros, que hemos denominado *Coincidencia mitótica (Mco).* Consiste en la frecuencia con que dos células vecinas en un tejido sólido se hallan simultáneamente en mitosis. El interés de esta medición reside en que dicho valor viene determinado, en parte, por la coincidencia mitótica de células hermanas, es decir, la frecuencia con que dos células procedentes de una misma partición celular se dividen simultáneamente en el siguiente ciclo. De acuerdo con la concepción clásica esto debería ocurrir con una frecuencia relativamente elevada, mientras que de acuerdo con el modelo de probabilidad de transición esto sería un suceso más bien raro, ya que las células hermanas, sincronizadas al terminar la división que las produce, perderían esta sincronía al pasar a través de la discontinuidad del *G*1 o estado *A*. En un trabajo publicado el pasado mes de mayo en *Nature* (J. Domingo, J. Serratosa, C. Vidal y E. Rius, *Nature* 273, 50-52, 1978) desarrollamos matemáticamente estos conceptos que nos llevan al cálculo de la coincidencia mitótica basándonos en el nuevo modelo o en el clásico. Comparamos los valores calculados con los observados y los resultados obtenidos son claramente favorables al modelo de probabilidad de transición. (J.D.)

Voces en la luz

El sistema telefónico de la Bell en los Estados Unidos se ha ido poco a poco desarrollando hasta poner a prueba la capacidad de transmitir mensajes por ondas luminosas, en lugar de hacerlo por ondas eléctricas y radioondas, que son las que ahora se usan. El principal motivo para tal esfuerzo es que la capacidad de un canal de comunicación dado aumenta con la frecuencia de la señal, y la frecuencia de las ondas luminosas en unas 1000 veces la frecuencia de las radioondas, más cortas. En 1976, el sistema Bell comenzó a experimentar una onda luminosa prototípica en Atlanta, y el año pasado abrió una red de ondas luminosas de 2400 metros en Chicago, bajo condiciones típicas de funcionamiento. Estos sistemas de prueba han tenido tanto éxito que la Bell quiere poner en marcha para finales de 1980 su primera red de comunicaciones por ondas luminosas.

Los resultados de las pruebas de Chicago están recogidos en los *Bell Laboratories Record*. La fuente luminosa es un láser de larga duración de arseniuro de aluminio galio o un diodo emisor de luz de alta radiación; en el sistema de Chicago son intercambiables. La luz se conduce por fibras de cristal muy transparentes. del grosor de un cabello. El cable óptico instalado para la red de funcionamiento contiene 96 de estas guías de luz. Un solo par de guías de luz puede soportar 672 conversaciones simultáneas de ida y vuelta, mientras que un par convencional de guías de cobre soporta 24.

Se ha demostrado que los láseres merecen toda la confianza. Hasta mediados de marzo el sistema sólo tuvo un fallo imputable al láser en más de 100.000 horas de operación acumulativa de todos los láseres. (Los laboratorios Bell han anunciado recientemente una nueva generación de pequeños láseres de estado sólido con una vida media probable de un millón de horas, es decir, unos 100 años.) Es más. durante los diez primeros meses había un margen de error de transmisión de menos de un segundo por día, lo que significa que el sistema está libre de errores el 99'999 por ciento del tiempo (medido en intervalos de un segundo). Otra medida de fiabilidad es la cantidad de tiempo que un sistema de telecomunicaciones pueda estar estropeado. El sistema de transmisión habitual de corta distancia está diseñado para no fallar más de dos horas al año. El sistema de ondas luminosas de Chicago tuvo un único fallo de 20 segundos en los diez primeros meses. Los resultados de la

prueba de Chicago, según el *Record,* confirman la esperanza de que la comunicación por ondas luminosas "tiene la capacidad para proporcionar ventajas muy notables en el campo de las telecomunicaciones".

Microscopía por ultrasonido

Los avances en la tecnología de imágenes ultrasónicas, que empezaron hace más o menos una década con aplicaciones macroscópicas, sobre todo en la ciencia de los materiales y en investigación biomédica, están, al parecer, a punto de proporcionar un nuevo instrumento de múltiple uso: un microscopio acústico de onda ultracorta con un poder de resolución comparable al de los microscopios ópticos existentes. Ya está en el mercado un tipo de microscopio acústico, el Sonomicroscopio, desarrollado y fabricado por Sonoscan Inc. Con una frecuencia de 100 megahertz (cien millones de ciclos por segundo), tiene un poder de resolución de 20 micrometros, y se dice que puede operar a frecuencias más altas, y, por tanto, con resoluciones más elevadas.

Actualmente se están haciendo investigaciones sobre tipos de microscopios acústicos más avanzados. Calvin F. Quate, de la Universidad de Stanford, encabeza uno de los grupos experimentales más importantes que está trabajando sobre el problema de aumentar la frecuencia operativa, y, por tanto, el poder de resolución de dichos aparatos. En un artículo reciente de *Mosaic*, publicación de la National Science Foundation, se cita a Quate refiriendo que, dentro de un año, su grupo espera obtener ya fotografías hechas con ondas de sonido a "longitudes de onda iguales a las longitudes ópticas", lo que correspondería a una frecuencia acústica de unos tres gigahertz (tres mil millones de ciclos por segundo). "Cuando este instrumento funcione, añade Quate, todos los detalles que ahora se consiguen en micrografías ópticas se lograrán también en micrografías acústicas."

Los esfuerzos del grupo de Stanford para desarrollar un microscopio acústico satisfactorio empezaron en los años 60 tras sugerencia de Rudolf Kompfner. El aparato de Stanford en su forma simple, diseñado por Quate en colaboración con Ross A. Lemons y otros, se describe en el artículo de *Mosaic* como un "instrumento de barrido que mueve la muestra a través de un haz acústico enfocado. Dos varillas cortas de zafiro forman los elementos claves del instrumento. Sirven para enfocar el haz en un líquido celular que sostiene la muestra, y para recoger la energía acústica repartida por el objeto. A cada uno de los extremos de cada varilla de zafiro va fijada

una película piezoeléctrica de óxido de cinc, mientras que la cara más interior de cada varilla va conectada a una lente cóncava esférica. La energía eléctrica se convierte en energía acústica mediante la película piezoeléctrica, se transmite a lo largo de la varilla hasta la lente, que la enfoca sobre la muestra, situada en el líquido celular entre los extremos de la varilla cóncava por tensión superficial. Las superficies cóncavas actúan como lentes positivas enfocando el haz de sonido de entrada en el líquido hacia un anillo estrecho originado en los puntos focales, donde se encuentran las lentes de entrada y salida... En teoría, el anillo puede menguar y menguar, enfocando más y más, mediante la simple consecución de frecuencias de sonido más y más elevadas".

Debido a que el campo de visión de este haz altamente enfocado es pequeño y despreciable, continúa el artículo, "el objeto... se examina mecánicamente moviéndolo a través del anillo del haz. La imagen examinada se construye y registra línea por línea y se exhibe en una pantalla de televisión". Según Quate, el inconveniente mayor de dicho sistema es que "lleva tiempo examinar toda una especie... Si es un objeto estático, esto puede no importar, pero si se trata de una especie viva y se mueve, ello podría ocasionar algún problema. Sin embargo, la duración del examen es lo bastante corta como para que podamos conseguir buenas imágenes de células vivas"

Las aplicaciones reales y potenciales de los microscopios acústicos de alta resolución varían desde análisis de materiales a escala mínima hasta el estudio de tejidos y órganos biológicos. Según Quate, la ventaja principal de la microscopía acústica sobre la microscopía óptica, para tales usos, proviene del hecho que, mientras el microscopio óptico depende –para su función de imágenes– de la distribución de los índices de refracción del objeto, el microscopio acústico "depende para su contraste de la distribución de las cualidades mecánicas del objeto: esto es, su densidad, elasticidad y viscosidad". Estas propiedades son, con mucho, más importantes para la función del tejido vivo que el índice de refracción óptico".

Poco antes de su muerte Kompfner señaló: "el microscopio acústico es un instrumento importante que con el tiempo se situará a la altura del microscopio electrónico y el microscopio óptico para examinar casi cualquier cosa que pueda ser objeto de estudio microscópico... Este aserto se funda no tanto en la resolución que ahora podemos obtener, que es bastante apreciable, sino más bien en el hecho de que estamos observando propiedades de la materia diferentes".

Imágenes nocturnas de la Tierra vista desde el espacio

Las fotografías registradas a medianoche por los satélites revelan un aspecto insólito de la Tierra. Las luces que se observan en su cara obscura son las llamas que se producen por la combustión del gas residual de los campos petrolíferos

Thomas A. Croft

Cuando la Tierra se observa desde el espacio durante el día, las nubes y los perfiles geográficos dominan el paisaje. La impronta de la presencia del hombre, cuando llega a ser visible, tan sólo modifica muy ligeramente el aspecto de estas imágenes. Sin embargo, durante la noche, y especialmente a medianoche en las noches sin luna, los destellos de las luces brillantes originadas por el hombre destacan sobre el fondo obscuro de la Tierra. Además de las luces ocasionales de los fenómenos naturales, tales como los relámpagos, los incendios forestales y la aurora, la cara obscura de la Tierra aparece salpicada por las luces de las ciudades y los fuegos de origen agrícola. Sin embargo, con mucho, las luces más brillantes de origen artificial son las llamas producidas por la combustión del gas residual de los grandes campos petrolíferos del mundo.

Este aspecto inusitado de la Tierra se ha puesto de manifiesto muy recientemente como resultado del Programa de Satélites Meteorológicos para la Defensa de las Fuerzas Aéreas de los Estados Unidos. La misión fundamental de los ingenios espaciales puestos en órbita en el curso de este programa consiste en proporcionar las imágenes que sirven de base para predecir la nubosidad a corto plazo. En un principio, las imágenes transmitidas a Tierra con este fin eran desechadas rutinariamente tras un día de uso, pero muy pronto se advirtió la pérdida de datos científicos en potencia que esto traía consigo y se tomaron las medidas oportunas para conservar las mejores imágenes de las auroras. Estas fotografías se hicieron públicas finalmente en el año 1973. A partir de entonces, se creó un servicio dedicado a archivar las fotografías originales en la Universidad de Wisconsin y distribuir las imágenes de las auroras en forma de microfilms a todos los investigadores interesados por ellas a través del Departamento de Comercio. La mayoría de las fotografías que se incluyen en este artículo proceden así del archivo de Wisconsin.

Características del sistema

El satélite meteorológico de las Fuerzas Aéreas no es el único sistema capaz de proporcionar imágenes nocturnas de la Tierra desde el espacio. Los tres ingenios espaciales Landsat, lanzados por la Administración Nacional de la Aeronaútica y el Espacio (NASA) en 1972, 1975 y 1978, podían operar también por la noche, aunque no fueron diseñados con este propósito y se han utilizado muy raramente en la toma de imágenes nocturnas. El sistema Landsat ofrece la ventaja de proporcionar fotografías en cuatro colores; por otro lado, su poder de resolución es mayor que el de los satélites meteorológicos de las Fuerzas Aéreas. Sin embargo, este último sistema es mucho más sensible a la captación de fuentes luminosas débiles. En el estudio de las fuentes luminosas más brillantes, los dos sistemas son complementarios: el satélite de las Fuerzas Aéreas resulta más adecuado para llevar a cabo un rastreo de larga duración sobre la Tierra, en tanto que el sistema Landsat puede proporcionar fotografías en color, de alta resolución, de zonas específicas previamente seleccionadas.

Los ingenios espaciales de estos dos programas norteamericanos se mueven en órbitas prácticamente circulares a alturas del orden de 800 o 900 kilómetros. Estas órbitas se eligen de suerte que pasen a unos nueve grados de los polos terrestres, haciendo que el plano de las órbitas efectúe un movimiento de precesión alrededor de la Tierra con la misma velocidad con que ésta se mueve alrededor del Sol. En relación con dichas misiones fotográficas, estas órbitas sincrónicas con el Sol presentan una ventaja muy importante: la hora del día (o de la noche) bajo la cámara no cambia en el curso de las estaciones y, en consecuencia, el cambio de la correspondiente iluminación es mínimo. Como resultado de este plan de vuelo, las imágenes que proporcionan los satélites quedan registradas en forma de tiras sucesivas orientadas aproximadamente en la dirección norte-sur; cada una de esas tiras corresponde a una órbita diferente del satélite. La hora del día en que se fotografía cada zona permanece constante al pasar de una tira a otra. El solapamiento entre dos tiras consecutivas aumenta para las latitudes más elevadas: los polos, situados a 90 grados de latitud Norte y Sur, aparecen en todas las tiras; los polos se fotografían alternativamente cada 102 minutos.

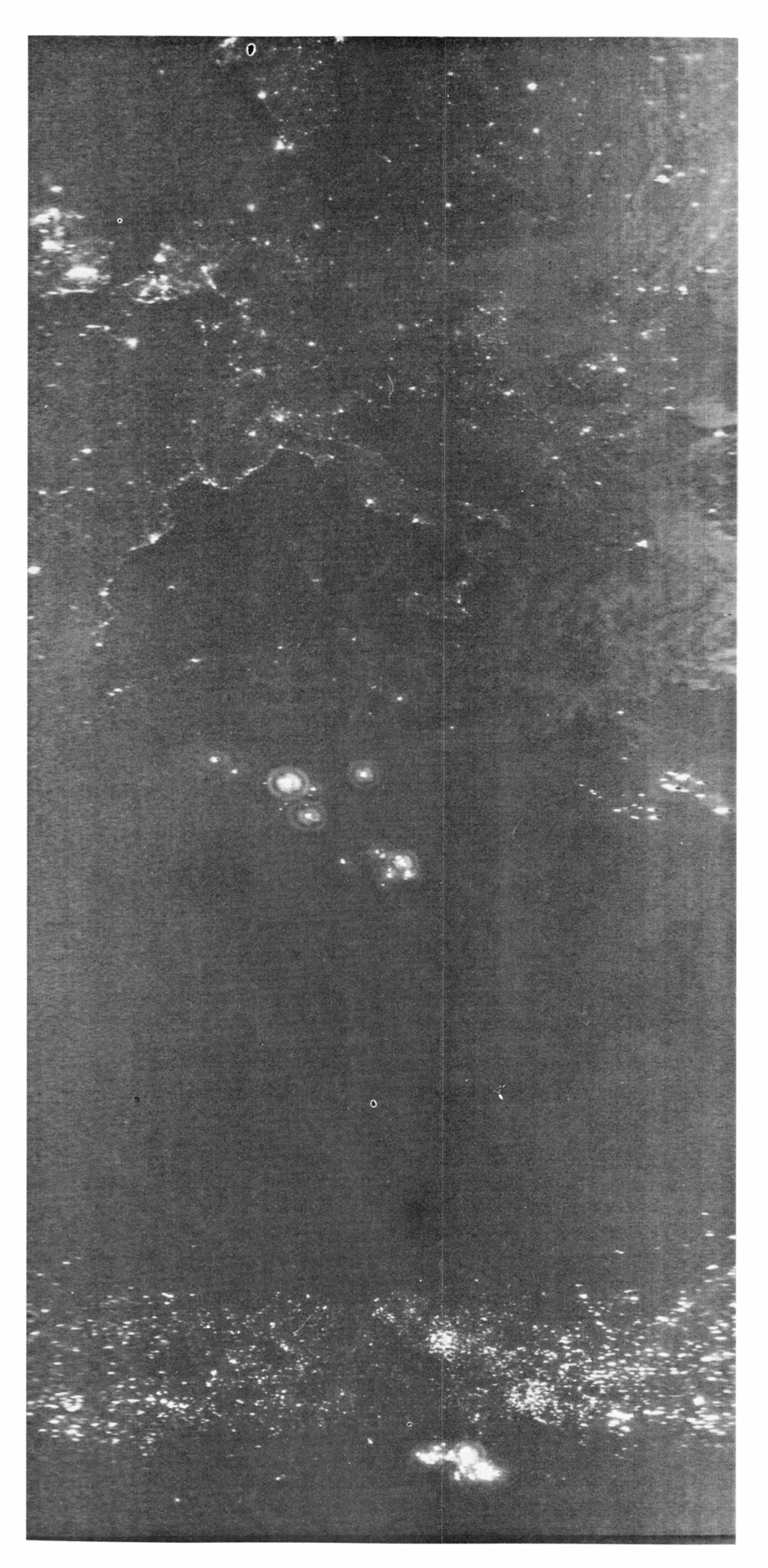
Uno de los hechos más sorprendentes que se ponen de manifiesto en muchas de las imágenes obtenidas con este sistema es la fuente luminosa: la luz de la Luna a medianoche. En su otra semiórbita, el satélite obtiene una imagen de la Tierra a la luz del día similar a ésta, pero mucho más detallada. El programa del satélite proporciona imágenes de las nubes que cubren el cielo sobre cualquier lugar de la Tierra cada 12 horas aproximadamente, durante los períodos en que la luz de la Luna ilumina el lado obscuro de la Tierra (es decir, aproximadamente durante la mitad de todo el tiempo). A bordo de cada uno de los satélites citados hay también otros tipos de sensores. Gracias a un dispositivo adecuado se toman simultáneamente imágenes infrarrojas, basadas en la radiación térmica procedente de la Tierra y de las nubes. Las películas resultantes muestran las temperaturas relativas de las diferentes nubes y permiten obtener un mapa de la nubosidad que cubre la Tierra incluso en las noches sin luna.

En las imágenes tomadas a plena luz lunar, cuando las cámaras del satélite operan a sensibilidad reducida, se hacen visibles

muchas luces y fuegos producidos por el hombre, pero tan sólo las llamas de gas son lo suficientemente brillantes como para ponerse de manifiesto con absoluta claridad *véanse, las ilusraciones de las dos páginas siguientes).* Sin embargo, a medida que disminuye la luz de la Luna, las fotografías sufren una transformación, de forma análoga al cambio que tiene lugar en el cielo al atardecer, cuando desaparece el sol y surgen las estrellas, ¡con la diferencia de que ahora lo que surge son las ciudades y los fuegos! Atendamos, por ejemplo, a la ilustración de la derecha obtenida por un satélite. En el norte, la bota de Italia se destaca claramente a la luz de la Luna y es fácil localizar muchas de las grandes ciudades de Europa. Los puntos más brillantes de esta fotografía corresponden a llamas de gas producidas en los campos petrolíferos de Argelia, Libia y Nigeria. El conjunto prácticamente uniforme de pequeñas luces que se observan a lo largo de Africa, al sur del Sáhara, parece estar originado por los incendios producidos deliberadamente para limpiar la tierra destinada a usos agrícolas o ganaderos. Imágenes de fuegos similares a éstas se ven por todo lo ancho del mundo.

En esta misma fotografía se pueden apreciar también algunas de las limitaciones que caracterizan al sistema del satélite meteorológico de las Fuerzas Aéreas. Los detalles situados cerca de los bordes de la imagen aparecen bastante borrosos, debido a la baja altura a que se mueve el satélite (su altura de vuelo es aproximadamente igual a la tercera parte de la anchura de esta imagen). El satélite meteorológico de las Fuerzas Aéreas que está en uso actualmente ha sido dotado de un sistema de observación de nuevo diseño cuyas imágenes presentan una mejor resolución cerca de los bordes.

El sistema con el que se obtuvieron estas fotografías trabaja de la forma siguiente. La imagen se origina en un telescopio óptico que barre la superficie de la Tierra en una dirección perpendicular a la trayectoria con resolución angular constante (en las regiones no polares, la dirección del barrido es aproximadamente este-



TRES FUENTES luminosas principales asociadas con las actividades humanas se observan en esta imagen, tomada recientemente en el curso del Programa de Satélites Meteorológicos para la Defensa de las Fuerzas Aéreas. Los numerosos puntos brillantes repartidos por el tercio superior de la figura corresponden a las luces de las ciudades de Europa. Las luces más grandes que se observan en el centro y la parte inferior de la imagen se deben a las llamas de gas de los campos petrolíferos de Argelia, Libia y Nigeria. (Los anillos que rodean las luces más brillantes están producidos por las reflexiones internas que tienen lugar en el interior del telescopio del satélite.) El conjunto uniforme de luces más pequeñas dispersas a lo largo de Africa, al sur del Sáhara, refleja los incendios de tierras de cultivo y de pastos.

oeste). En el plano focal del telescopio existe un sensor óptico; este instrumento es similar al fotómetro de un fotógrafo excepto en el hecho de que su campo de visión abarca tan sólo dos décimas de grado. Los extremos de la imagen están muy alejados del satélite y se ven oblicuamente, de forma tal que el campo de visión angular del telescopio cubre una superficie seis o siete veces mayor que la superficie que cubre en la dirección del nadir (es decir, con la cámara dirigida recta hacia abajo). Los datos que proporciona este sistema son enviados a tierra en forma de un flujo de números binarios que representan las variaciones de brillo registradas por el sensor de barrido. Más tarde, en tierra, estos números se reconvierten en puntos luminosos y las correspondientes imágenes se recomponen sobre una película en forma de mapa. En este proceso se realizan algunas correcciones aproximadas de las distorsiones, que se producen por diversas causas. (La órbita no es exactamente circular, la Tierra no es exactamente esférica, el satélite no se encuentra siempre a la misma altura, etc.).

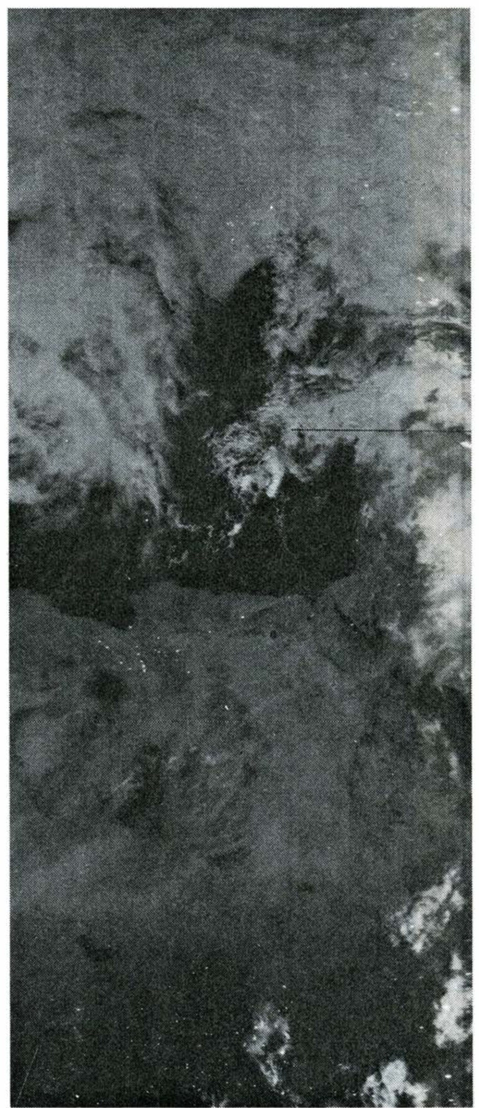
Algunas de las llamas de gas más grandes son tan brillantes que se hacen visibles las reflexiones internas producidas dentro del cuerpo del telescopio; este efecto es el responsable de los anillos que se observan alrededor de las llamas de gas en algunas de la imágenes. Algo similar sucede en los telescopios astronómicos, cuyo objetivo es también la detección de objetos muy pequeños, y muy brillantes, en relación con sus alrededores. Excepto en condiciones de máximo contraste, tal como el que existe entre las llamas brillantes y su entorno obscuro, este hecho no representa ningún problema en el sistema que produce las imágenes en el satélite.

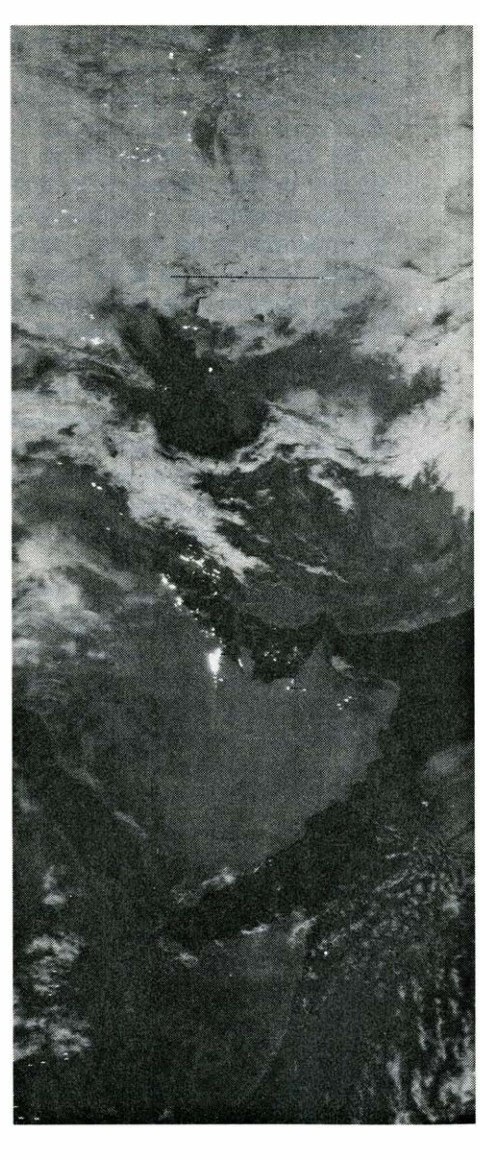
Si se procede a una comparación entre tales imágenes y sus equivalentes infrarrojas, cuyo tamaño es exactamente el mismo y cuyo ajuste se realiza con gran facilidad superponiendo las dos fotografías sobre una mesa luminosa, se puede ver que las luces de las ciudades atraviesan las nubes delgadas con poca degradación. Las nubes de espesor medio difunden la imagen y las nubes muy espesas pueden impedir totalmente el paso de la luz.

Llamas de gas

Las imágenes nocturnas de los satélites ponen de manifiesto la existencia de llamas de gas brillantes en muchas partes del mundo, si bien la mayor concentración de ellas se presenta en la región del Golfo Pérsico. Aun a plena luz lunar, esta zona parece arder bajo la presencia de muchas llamas de gran tamaño *(véase la ilustración inferior de la página 46)*. La combustión del gas residual de los campos petrolíferos







VISION PANORAMICA a la luz de la Luna de la zona terrestre comprendida entre el noroeste de Africa y el sudeste de Asia. La figura está compuesta por fragmentos de seis imágenes en forma de bandas, obtenidas por el satélite meteorológico de las Fuerzas Aéreas a lo largo de seis órbitas sucesivas en la dirección norte-sur. Todas las tomas fueron realizadas en la medianoche local del 6 de febrero de 1974. La iluminación principal se debe a la luz reflejada por

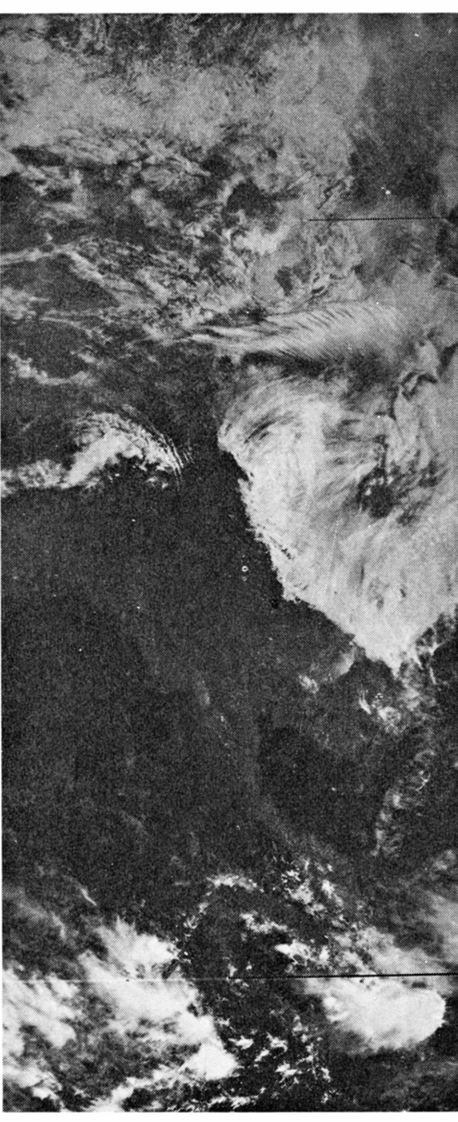
del Golfo Pérsico ha constituido uno de los espectáculos más grandiosos de esta parte del mundo durante muchas décadas y ya fue señalado por los aviadores que sobrevolaron la región en 1933. La práctica de dar salida a la atmósfera al gas de desecho de los pozos de petróleo, sin quemarlo, data de tiempos todavía más lejanos; hace más de 100 años un artículo de *Scientific American* citaba este hecho en relación con los primeros campos de petróleo de Pennsylvania; el autor estimaba que uno solo de estos pozos descargaba 25.000 metros cúbicos de gas por hora y que lo había estado haciendo así "durante años y años".

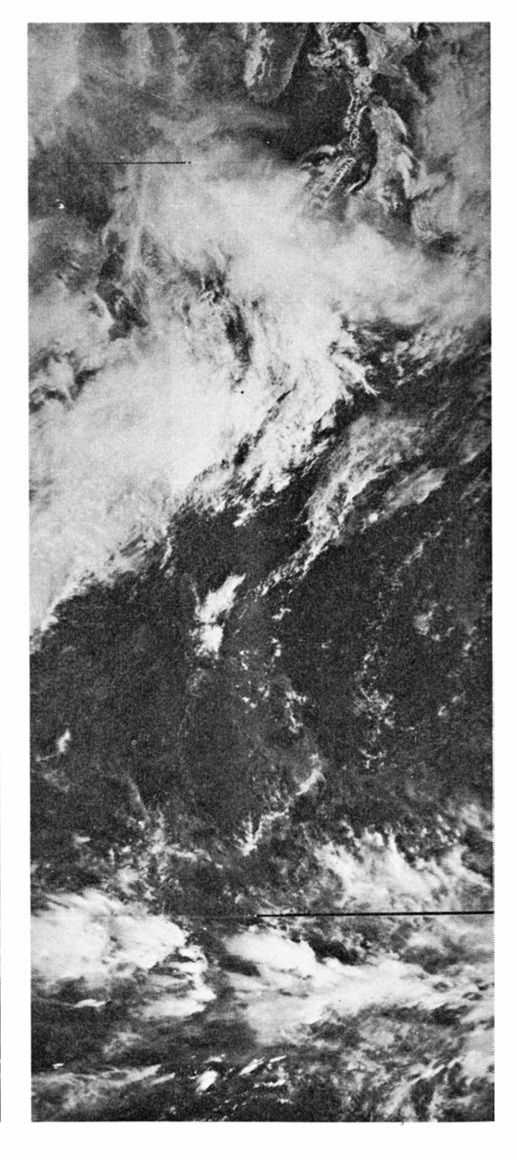
Puesto que los hechos más destacados de estas imágenes nocturnas son las llamas de gas, vamos a dedicarles una mayor atención. El sistema de observación del satélite meteorológico de las Fuerzas Aéreas se ha diseñado para medir solamente el brillo medio de cada uno de los elementos de sus imágenes, llamados "pixels", cuyo lado mide aproximadamente tres kilómetros de longitud. (El tamaño exacto varía de uno a otro lugar de la imagen y de sensor a sensor.) En consecuencia, la contribución de una simple llama a este brillo medio debe depender no sólo de su brillo intrínseco sino también del porcentaje de la superficie del pixel que esté ocupado por la imagen de la llama. En otras palabras, las llamas deben ser o muy brillantes o muy grandes para que sus imágenes se pongan de manifiesto con claridad. Lo que sucede es que las llamas de gas en cuestión cumplen ambos requisitos. Dichas llamas cubren una gran superficie como consecuencia de la práctica de dispersar el fuego en sentido lateral con el fin de asegurar su combustión total. En general, esto se lleva a cabo suministrando el gas a través de una serie de conducciones dispuestas en una fila cuya longitud es al menos de 50 metros *(véase la ilustración superior de la página 47).*

En un esfuerzo por analizar con más detenimiento dos de las llamas mayores de Argelia –que se pueden observar en la figura de la página 43– obtuve las fotografías diurnas proporcionadas por el Landsat correspondientes a esta misma re*gina 47).* Las dos llamas en cuestión se ven con claridad en varias de estas fotogra-
ven con claridad en varias de estas fotografías, con una separación norte-sur de unos 11 kilómetros. En términos del sistema de observación del satélite de las Fuerzas Aéreas, las dos llamas están separadas entre sí tres pixels.

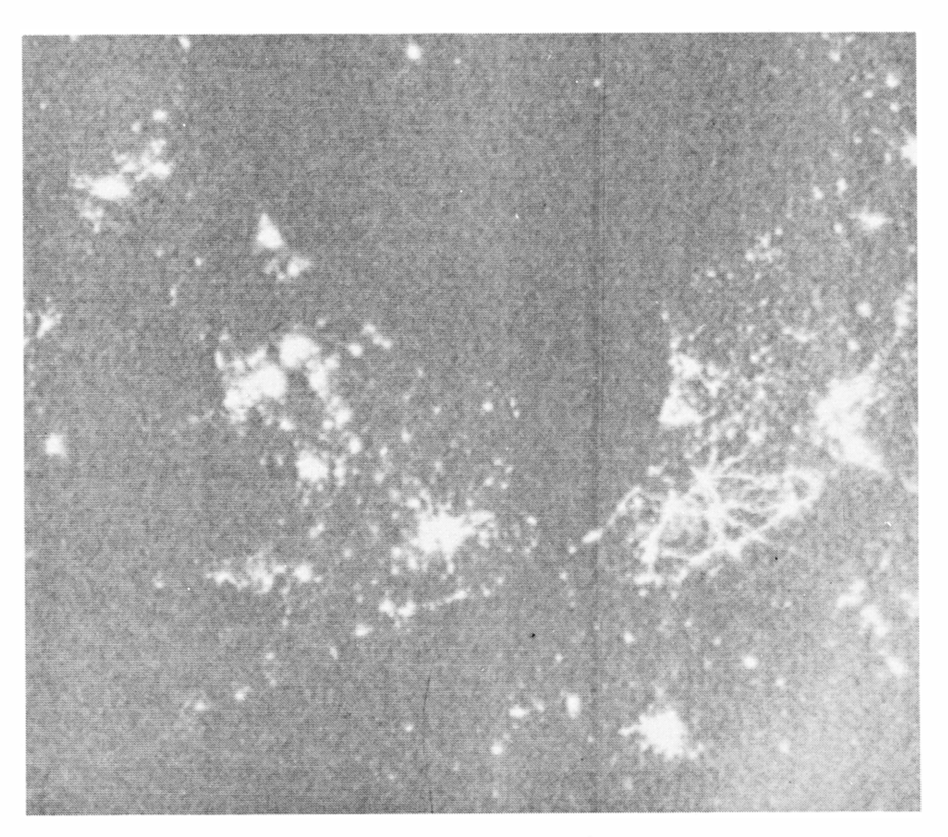
Observando con detenimiento las transparencias ampliadas de las fotografías del Landsat con ayuda de una lupa, logré determinar que las dimensiones de las llamas rojas son del orden de 500 metros, es decir, estas llamas son mucho mayores que



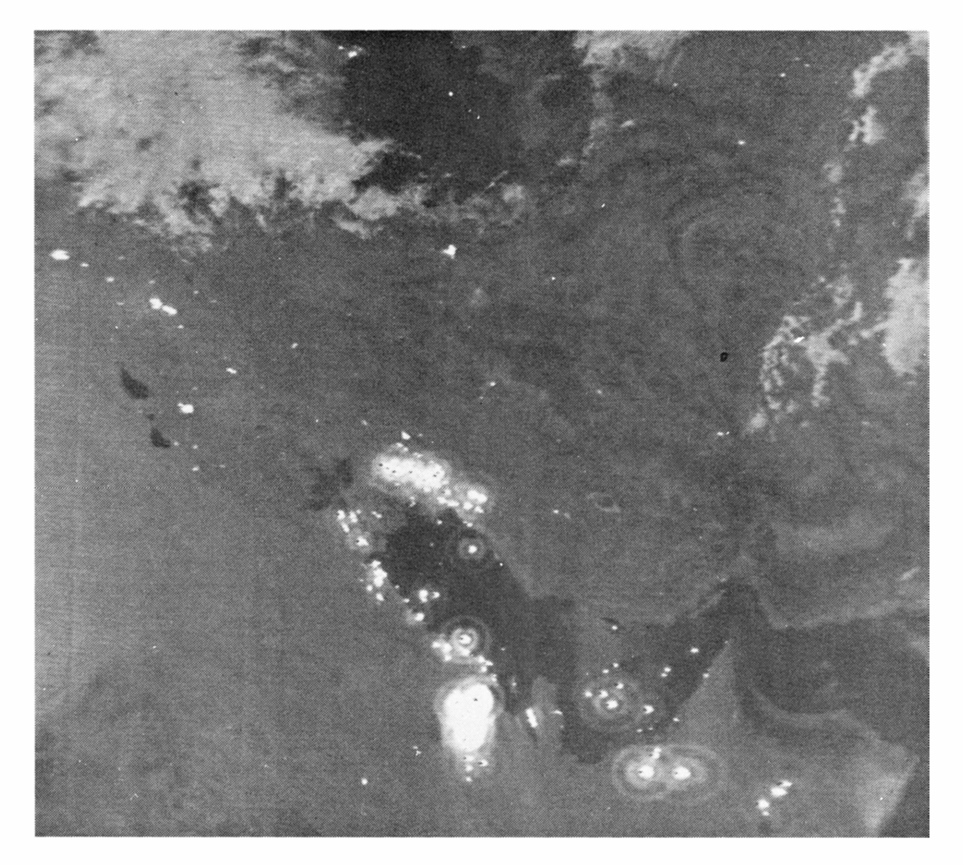




la Luna. El solapamiento entre las bandas adyacentes aumenta para grandes latitudes. En este caso, las bandas fueron recortadas por arriba y por abajo; los originales abarcan los dos polos en cada banda. Aunque se destacan algunas luces de ciudades y fuegos agrícolas sobre el conjunto brillantemente iluminado por la luz de la Luna, sólo las llamas producidas por la combustión de gas son lo suficientemente luminosas como para ponerse de manifiesto con claridad.



LAS LUCES DE LAS CIUDADES de los países que circundan el Canal de la Mancha y el mar del Norte se ponen de manifiesto con una claridad poco común en esta imagen ampliada de un satélite de las Fuerzas Aéreas estadounidenses, a pesar de su contraste relativamente pobre. La fotografía se tomó en una noche sin luna del mes de mayo de 1977, con los sensores del satélite operando a la máxima sensibilidad posible. Las Islas Británicas están a la izquierda. El conjunto de luces a la derecha del centro corresponde a Bélgica. El punto luminoso que se observa en la parte inferior derecha de la imagen, cerca del borde, es París.



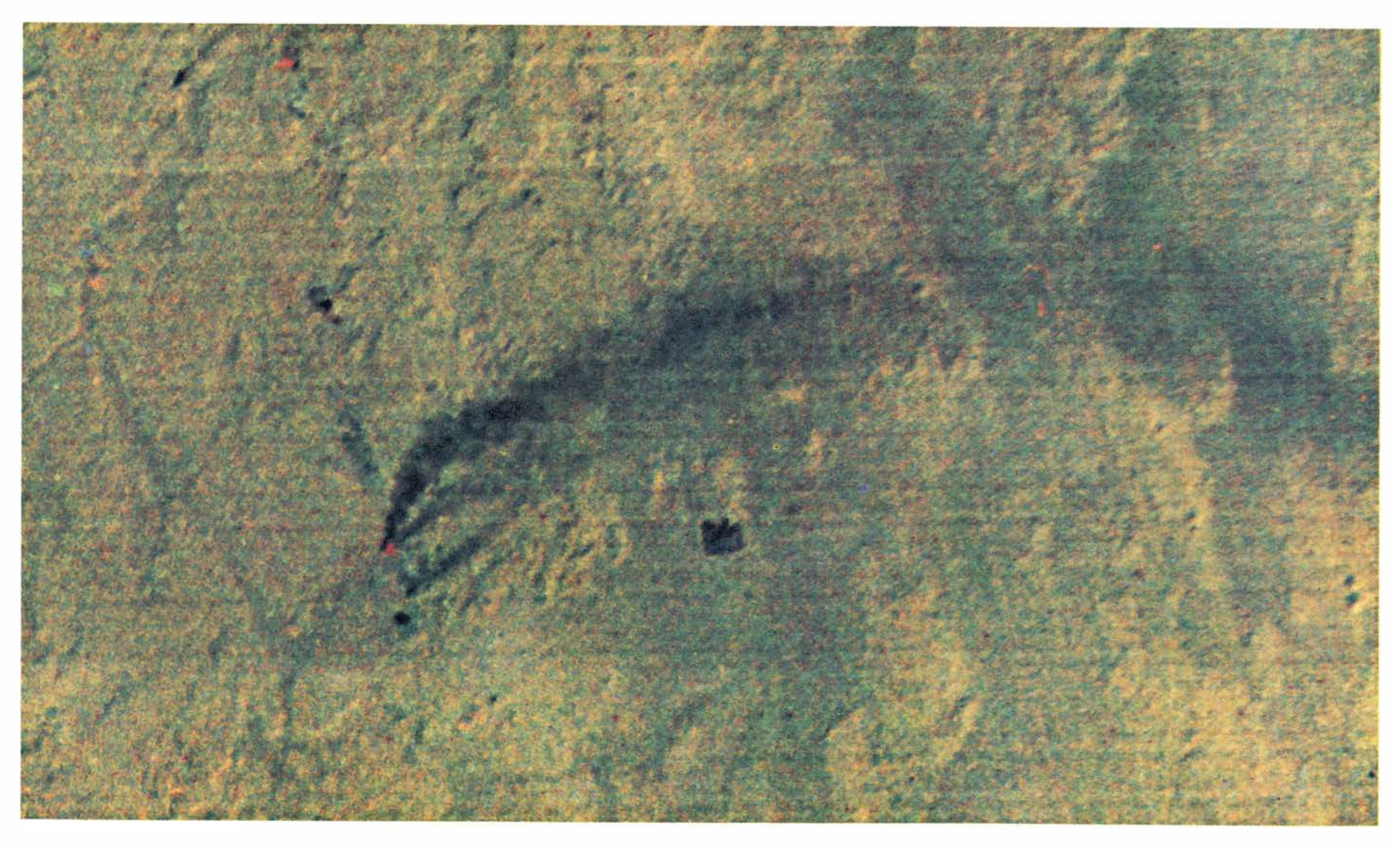
LAS LLAMAS DE GAS aparecen concentradas en la región del Golfo Pérsico, tal como muestra esta imagen ampliada, tomada a la luz de la Luna por el sistema óptico del satélite de las Fuerzas Aéreas. La zona obscura de la parte superior es el mar Caspio; el punto brillante situado en el promontorio es Baku, en la Unión Soviética; el punto mayor que aparece al sur del Caspio es Teherán, capital de Irán; finalmente, el punto similar al anterior, a la izquierda del centro de la imagen, corresponde a Bagdad, capital de Iraq.

las que se observan generalmente en las fotografías terrestres. Estas llamas de gas se detectan con facilidad en las imágenes diurnas de los satélites, puesto que el viento arrastra sus penachos de humo varias decenas de kilómetros. En un intento de evitar esta demostración gráfica de la correspondiente polución, los quemadores de gas más modernos están equipados con dispositivos de combustión "sin humo". Esta práctica está justificada indudablemente desde el punto de vista de las apariencias, pero dada la preocupación que existe hoy en día acerca del nivel creciente de dióxido de carbono en la atmósfera, desde una perspectiva a largo plazo podría ser más ventajoso para la humanidad el hecho de que no se quemaran algunas de las partículas de carbón existentes en el humo.

Una estimación más precisa de las dimensiones laterales de estas llamas se puede obtener a partir de las imágenes nocturnas del Landsat. A petición del Departamento de Geología de los Estados Unidos, uno de los satélites Landsat fue activado con este fin cuando volaba sobre Argelia en la noche del 31 de enero de 1976. En la ilustración superior de la página 48 se reproduce la máxima ampliación posible de una de las llamas fotografiadas en esta ocasión. Se trata de la llama mayor, y situada más al sur, en la fotografía del Landsat que se reproduce en la parte inferior de la página opuesta.

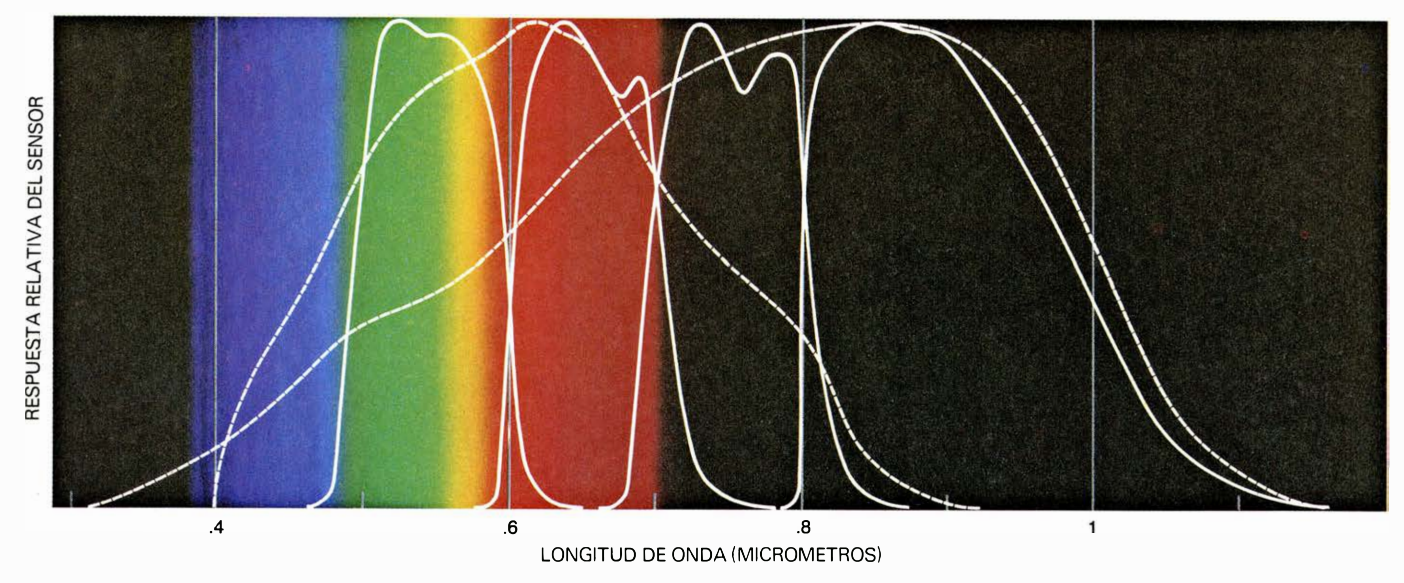
Para comprender lo que esta imagen nocturna en particular pone de manifiesto, es necesario conocer algo más acerca de cómo se obtuvo. Cuando el satélite Landsat sobrevuela una región, su telescopio de barrido multiespectral se enfoca hacia abajo por medio de un espejo de 45 grados que gira hacia atrás y hacia adelante con el fin de barrer toda la escena de izquierda a derecha (es decir, en una dirección perpendicular a la línea de vuelo). Los rayos luminosos que llegan al plano focal del telescopio se dirigen hacia unos sensores electrónicos a través de una serie de minúsculas conducciones ópticas de vidrio, cada una de las cuales tiene una sección transversal que corresponde a un cuadrado de 79 metros de lado sobre la superficie de la Tierra en la dirección del nadir. (En el borde de la imagen, a 93 kilómetros de la proyección de la trayectoria del satélite sobre la Tierra, cada conducción óptica cubre una superficie algo mayor, de modo que el poder de resolución medio de sistema se supone igual a unos 80 metros.) Dado que los sensores electrónicos se "leen" una vez cada 57 metros de barrido, los pixels del Landsat son rectángulos de 57 metros por 79 metros en el nadir. Un objeto brillante de grandes dimensiones puede afectar a dos o tres lecturas consecutivas del sensor, debido en parte a que las conducciones óp-

LLAMAS DE GAS DISTRIBUIDAS a lo largo de superficies muy extensas con el fin de asegurar que la correspondiente combustión sea completa. La presente fotografía, tomada desde un lugar adecuado situado en un equipo de perforación próximo, corresponde a un campo petrolífero de Kuwait. Las conducciones utilizadas para el suministro del gas tienen una altura de unos 2,5 metros. Esta fotografía fue tomada por David F. Cupp *(National Geographic.)*



CUATRICROMIA correspondiente a dos grandes llamas de gas existentes en Argelia, tomada a la luz del día por el satélite *Landsat 1*, operado por la Administracion Nacional de la Aeronaútica y el Espacio (NASA). Las llamas del gas en combustión tienen el aspecto de manchas rojas situadas a la cabeza de largas columnas de humo. En la parte superior de la página siguiente se reproduce una imagen ampliada de la llama situada más al sur de las dos.

AMPLIACION de una porción extremadamente pequeña de una imagen nocturna del Landsat conteniendo una de las llamas de gas argelinas que se reproducen en la ilustración inferior de la página precedente. Esta ampliación fue realizada por el autor del presente artículo con el fin de intentar evaluar la extensión lateral de la llama. Cada una de las imágenes del Landsat está compuesta por millones de pequeños elementos, llamados pixels, que corresponden a otras tantas zonas rectangulares de la superficie de la Tierra de 79 metros de longitud y 57 metros de anchura. Para cada pixel, el sistema que proporciona las imágenes del Landsat obtiene cuatro lecturas independientes del sensor, cada una de las cuales corresponde a un intervalo espectral diferente. Dado que en esta imagen ampliada sólo existían 35 pixels iluminados, se podía conseguir la coloración de cada uno de los rectángulos individuales por métodos fotográficos ordinarios. Con el fin de producir un desplazamiento de las radiaciones invisibles del infrarrojo y del infrarrojo próximo presentes en los datos del Landsat hasta la región visible del espectro, se utilizó un sistema de reproducción conocido como método del "falso color". La imagen resultante sugiere que la longitud de la llama de gas era del orden de 500 metros, aunque cabe que fuese compuesta por dos llamas menores separadas entre sí unos 200 metros.

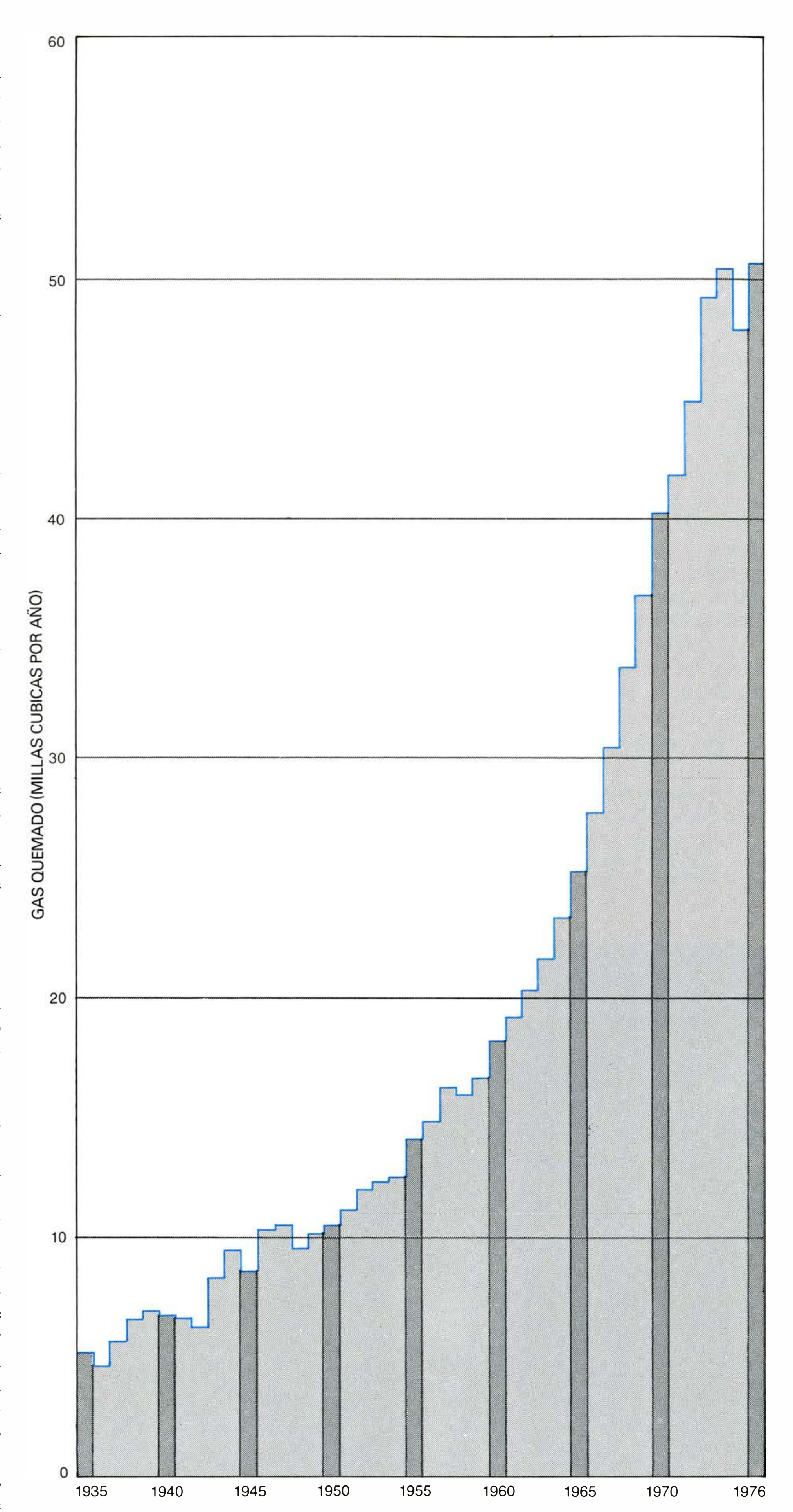


COMPARACION entre los intervalos espectrales del sistema de observación multiespectral del Landsat y las dos versiones del sensor de que están dotados los satélites meteorológicos de las Fuerzas Aéreas estadounidenses. Las seis curvas que se reproducen en esta figura tienen el mismo valor máximo, aunque en la actualidad tanto los viejos sensores (representados por la línea de trazos blanca de la derecha) como los nuevos sensores de las Fuerzas Aéreas (representados por la línea de trazos blanca de la izquierda) son más sensibles que los sensores del Landsat (representados por las cuatro curvas continuas blancas). La banda de colores que atraviesa el gráfico de arriba abajo para longitudes de onda comprendidas entre 0,4 y 0,7 micrometros corresponde a la región visible del aspectro electromagnético. La comparación entre las curvas de los sensores de los satélites para esta región del espectro demuestra por qué ambos ingenios espaciales son adecuados para la detección de fuegos: la energía de los fuegos se transmite para longitudes de onda a las que el ojo humano es insensible.

ticas son mayores que el pixel de 57 metros y en parte debido a un retardo en la electrónica. (Hace algunos meses, los delegados rusos de las Naciones Unidas sugerieron que no se debería permitir el uso de pixels menores de 50 metros; un tamaño inferior al citado sería considerado como una violación del derecho a la intimidad de una nación.)

Dentro de cada pixel, el sistema de observación del Landsat obtiene cuatro lecturas independientes del sensor, una para cada uno de otros tantos intervalos espectrales o colores. (El término "color" se utiliza aquí con cierta libertad; de hecho, dos de dichos intervalos se sitúan en las regiones espectrales del infrarrojo y del infrarojo próximo y, por tanto, son invisibles para el ojo humano.) Para cada color y para cada pixel el Landsat envía a tierra un número binario codificado en una forma tal que representa el nivel de potencia de la luz. Puesto que en la ampliación que estamos considerando aquí solamente hay 35 pixels iluminados, utilizando técnicas fotográficas ordinarias se pudo conseguir la coloración individual de todos los pequeños rectángulos que componen dicha imagen. Para hacer visibles los datos del infrarojo y del infrarojo próximo, adopté un método de reproducción de "falso color", desplazando todos los niveles digitales de potencia hacia el extremo de longitudes de onda más bajas del espectro en una cuantía de 0,2 micrometros. Procediendo de esta forma, el infrarrojo original se reproduce como rojo, el infrarrojo próximo como verde y el rojo como azul. (En la imagen de partida no existe prácticamente el verde.)

Una imagen típica del Landsat está constituida por 7.581.600 pixels, cada uno de ellos con cuatro lecturas espectrales independientes. De aquí se deduce que las imágenes mandadas desde el espacio en un solo día representan más de 30 millones de números. Ningún método fotográfico ordinario sería utilizable en la práctica para componer sobre una película la correspondiente imagen en color. De hecho, el sistema de que está dotado actualmente el Landsat ni siquiera utiliza luz. En su lugar, la película se mantiene al vacío y se expone al bombardeo directo por un haz de electrones. Este método permite exponer líneas muy finas sobre la película; cada una de estas líneas corresponde a una tira de pixels contiguos en una dirección perpendicular a la trayectoria. Las composiciones cromáticas se obtienen mecánicamente a continuación superponiendo las diferentes películas en blanco y negro de que se dispone, cada una de las cuales corresponde a una de las bandas originales del espectro. La reconstrucción de las imágenes del satélite de las Fuerzas Aéreas se



ESTIMACION ANUAL del volumen de gas (a presión atmosférica) quemado en los campos petrolíferos de todo el mundo entre 1935 y 1976. Los oportunos cálculos fueron llevados a cabo por Ralph M. Rotty, del Instituto para Análisis de la Energía de Oak Ridge. Haciendo uso de ciertas aproximaciones groseras, el autor de este artículo ha extrapolado estos datos hasta el año 1859, fecha del descubrimiento del petróleo, y hasta la actualidad. Se sabe que en poco más de un siglo se han perdido 4200 kilómetros cúbicos de gas.

lleva a cabo de forma similar, excepto en el hecho de que se utiliza un haz luminoso en lugar de un haz de electrones.

Una vez expuestas las bases del sistema de producción de imágenes del Landsat, vamos a reexaminar la fotografía superior de la página 48 e intentar visualizar la llama que produjo esta luz. Muchos de los pixels del borde de la imagen sólo muestran la existencia de una luz muy débil, que es la que cabría esperar si sólo una pequeña porción de llama estuviera presente en el correspondiente rectángulo. Aunque la longitud de toda la imagen es de unos 500 metros, la luz en cuestión podría deberse también a dos llamas separadas entre sí unos 200 metros. Por otra parte, el centro obscuro podría haber sido producido por el humo que oculta parte de la llama. (Las llamas a que nos estamos refiriendo no aparecen separadas en la imagen diurna del Landsat.) Quizás este problema se podría resolver definitivamente examinando más imágenes del Landsat de esta misma zona.

Al comparar las imágenes a gran escala de las Fuerzas Aéreas con los mapas publicados de los principales campos petrolíferos y de gas natural del mundo, se pone de manifiesto que muchas de tales llamas coinciden con los campos petrolíferos, pero pocas de ellas coinciden con los campos de gas. Este hecho no es sorprendente: un pozo de gas se puede clausurar con gran facilidad cuando el gas no se vende, en tanto que un pozo de petróleo libera gas inevitablemente siempre que se extrae petróleo. Frecuentemente, el petróleo llega a la superficie formando una mezcla espumosa con varios gases; el petróleo debe separarse de ellos antes de que pueda ser bombeado. Los gases en cuestión tienen que ser eliminados de alguna forma y en las regiones alejadas puede resultar más económico quemar el gas que transportarlo hasta encontrar un comprador. Los dirigentes gubernamentales deberían tener en cuenta el significado de estas fotografías al fijar el precio del gas natural. En efecto, podemos considerar que la depresión artificial de los precios del gas natural en los Estados Unidos y otros países desarrollados es la causa directa de este despilfarro en los campos petrolíferos de los países menos desarrollados. Presumiblemente, el dilema económico con que se enfrentan los dirigentes políticos en las economías controladas del mundo es similar al que acabamos de exponer aquí, puesto que las grandes llamas de gas se observan también en los campos petrolíferos de la Unión Soviética y otros países comunistas.

Puesto que la humanidad ha estado lanzando a la atmósfera o quemando este gas desde hace más de 100 años, es lógico que nos preguntemos a cuánto ascienden las correspondientes pérdidas. La respuesta es difícil, puesto que las partes más directamente involucradas en este asunto no están muy interesadas en ser demasiado francas acerca de los detalles. A pesar de estas dificultades, Ralph M. Rotty, del Instituto para el Análisis de la Energía de Oak Ridge, ha recopilado las estimaciones anuales del volumen total mundial de gas quemado desde 1935 a 1976 *(véase la ilustración de la página anterior).*

Estos resultados demuestran que la cantidad de gas quemado ha crecido de forma tal que supera el ritmo de crecimiento de un simple modelo exponencial, al menos con anterioridad al reciente aumento del precio del petróleo. Sumando todas estas estimaciones anuales, extrapolando las cifras hasta el año 1978 a un ritmo constante y retrocediendo hasta el año 1859, en que se descubrió el petróleo a un ritmo excepcional, he deducido que el volumen (a presión atmosférica) de gas desperdiciado es igual a 4200 kilómetros cúbicos. Rotty estima que en los últimos años las llamas de gas suponen un 3 por ciento del total de los hidrocarburos quemados por el hombre. Las observaciones nocturnas de los satélites que he descrito en el presente artículo podrían proporcionar una forma alternativa de calcular el ritmo con que dicho gas está siendo quemado actualmente en el mundo.



ESTA LLAMA DE GAS GIGANTE se observó en una de las imágenes tomadas por un satélite en enero de 1975. La llama, una de las más grandes que el autor ha detectado a lo largo de todo su trabajo sobre las fotografías obtenidas por las Fuerzas Aéreas, estaba localizada cerca de Surgut, al norte de Siberia. Al parecer, el campo petrolífero carece de un gaseoducto capaz de transportar el gas en cuestión hasta un lugar en que pudiera ser aprovechado. La línea del ferrocarril Transiberiano se puede seguir bien gracias a pequeñas y grandes ciudades que se suceden como cuentas de un collar en la parte inferior de la imagen.

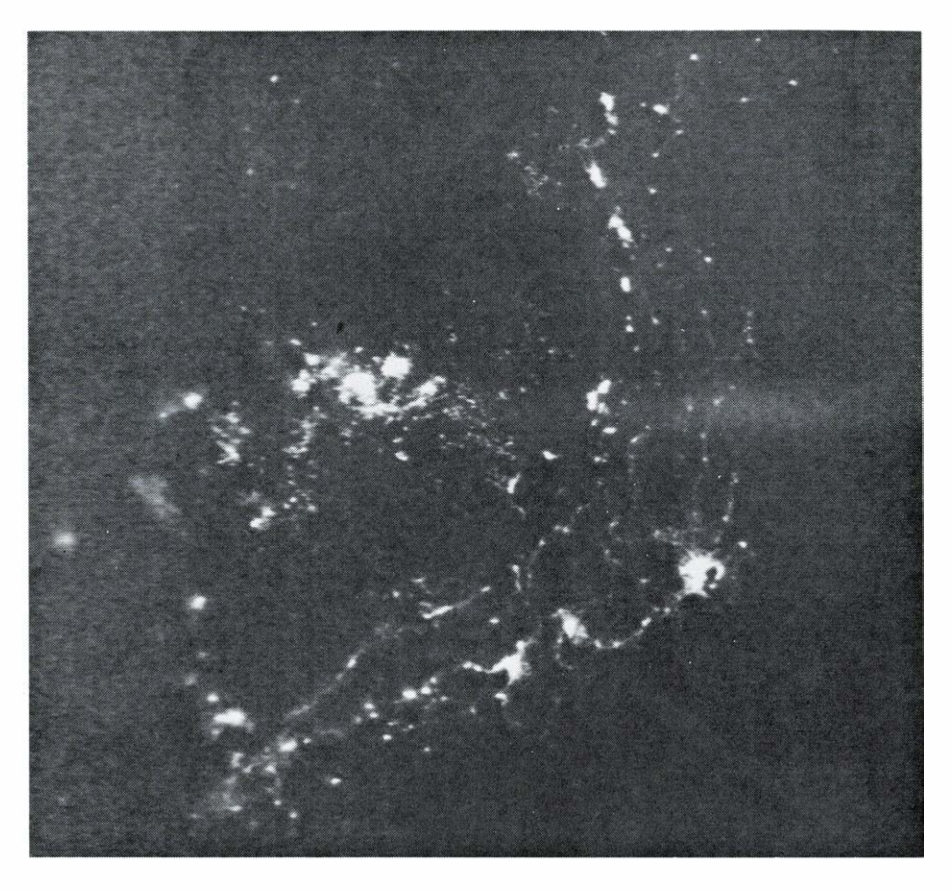
Linternas japonesas

Uno de los aspectos más curiosos del lado obscuro de la Tierra se puede contemplar en la fotografía tomada por un satélite que reproducimos en la parte superior de la página opuesta. A la derecha está el Japón, una de las naciones más luminosas del mundo por la noche, debido a su alta densidad de población y avanzado nivel de desarrollo. A la izquierda, en el centro del mar del Japón, una zona que aparece normalmente negra en las noches sin luna, se puede observar una gran concentración de luces brillantes. Las luces parecen estar dispuestas de forma ordenada y su configuración varía de una noche a otra. Desde hace tiempo se sabe que esta región está sometida a grandes tormentas acompañadas por muchos relámpagos. Sin embargo, en respuesta a mi petición de informes, la Agencia Meteorológica Japonesa afirmó que la noche en que se registró esta imagen no hubieron tormentas importantes en el mar del Japón. Una segunda consulta proporcionó una respuesta más satisfactoria. Los científicos de la Oficina de Pesca del gobierno japonés me proporcionaron los mapas que mostraban la localización de la flota dedicada a la pesca del calamar en la fecha en cuestión. La concordancia entre los mapas y las imágenes de los satélites resultaba indiscutible.

Más tarde fui informado de que durante la temporada de pesca del calamar grandes flotas de barcos japoneses y co-

reanos operan en el mar del Japón. Solamente la flota japonesa está compuesta por unos 2600 barcos, cada uno de los cuales pesa entre 60 y 100 toneladas; cada barco lleva unas 50 lámparas incandescentes con una potencia media de 3500 vatios por lámpara. ¡Esto supone algo más de 400 millones de vatios de luz eléctrica!

Las lámparas se utilizan para atraer a los calamares a la superficie del agua, donde son recogidos con redes. Cada barco dispone sus luces en una doble fila situada unos dos o tres metros por encima de la cubierta. La mitad de las bombillas no disponen de pantallas y las que posee la otra mitad son de un tamaño muy pequeño. Por término medio sólo la mitad de la flota pesquera japonesa se dedica a faenar a un mismo tiempo y, en consecuencia, la luz total emitida sería sólo del orden de 200 millones de vatios. Aun así, la luz en aquella zona es muy intensa. Por otro lado, además de estos barcos, hay que tener en cuenta también la presencia en dichas aguas de la flota pesquera del calamar procedente de Corea, menos importante que la anterior, y otra flota japonesa que opera en aguas menos profundas, dedicada a la captura de otro pez: el "saury". (Otra flota japonesa dedicada a la pesca del calamar suele trabajar en las aguas de Nueva Zelanda, pero hasta ahora su presencia allí no se ha observado en las imágenes que proporcionan los satélites meteorológicos de las Fuerzas Aéreas.)



LUCES BRILLANTES sobre el mar del Japón: se observan a la izquierda de esta imagen, tomada por un satélite de las Fuerzas Aéreas en una noche sin luna de julio de 1975. Estas luces coinciden con las posiciones de la flota japonesa que se dedica a la pesca del calamar. Los varios miles de barcos que constituyen dicha flota están dotados de bombillas incandescentes que sirven para atraer a los calamares a la superficie del mar y cuya potencia total asciende a centenares de millones de vatios. El propio Japón aparece nítidamente delimitado por las luces de las ciudades de sus costas, que se hallan densamente pobladas.

Se puede obtener una idea más clara de lo que representa esta potencia luminosa comparándola con la potencia total emitida hacia el cielo por toda la población de los Estados Unidos en forma de luz visible.

La cuantía de esta emisión fue medida hace una década y se encontró que su valor era del orden de unos 40 millones de vatios (es decir, aproximadamente 0,2 vatios por habitante). Teniendo en cuenta la baja eficiencia de las lámparas incandescentes que utilizan los pescadores en relación con las fuentes luminosas utilizadas en los Estados Unidos por las noches, he estimado que la suma de la luz de los Estados Unidos detectada por el satélite de las Fuerzas Aéreas es sólo dos o tres veces mayor que la procedente de la flota de pesca del calamar en el mar del Japón.

Evidentemente, las luces de las ciudades de cualquier país, incluyendo los Estados Unidos, se presentan mucho más dispersas. De hecho, su imagen refleja con bastante precisión la distribución de la población. Esta correspondencia puede apreciarse con facilidad comparando el montaje de las tres imágenes nocturnas de los Estados Unidos que se reproducen en la parte superior de la página siguiente con el mapa del censo de población proporcionado por un ordenador y que se reproduce en la parte inferior de la misma página. Tan sólo se ponen de manifiesto algunas discrepancias; por ejemplo, al este de Salt Lake City aparecen dos luces que no muestran la correspondiente contrapartida en el mapa del censo.

Incendios de origen natural y agrícola

En muchas regiones agrícolas la materia vegetal se quema regularmente como parte del ciclo estacional de cultivo. Anteriormente me he referido ya a los incendios de origen agrícola que se observan a lo largo del Africa central, al sur del Sáhara. En la fotografía de la página 53, se recoge otro ejemplo de este hecho captado cerca de Estados Unidos. Puesto que las luces que se observan a lo largo de la costa occidental de México están ausentes en otras épocas del año, cabe suponer que están originadas también por incendios agrícolas. Por otra parte, las llamas que muestra esa imagen varían poco en el transcurso del año. Las llamas de gas brillante que se observan en dicha foto están asociadas al campo petrolífero Reforma. situado en la costa oriental de México. (Esta foto en particular ilustra también perfectamente cuán al sur puede aparecer la aurora en algunas ocasiones.)

En otras muchas regiones del mundo se pueden ver incendios de origen agrícola similares a los anteriores. En general, el origen agrícola o forrajero de estos incendios se puede deducir del hecho de que desaparecen en ciertas épocas del año. Las imágenes diurnas del Landsat permiten estudiar la amplia banda de combustiones agrícolas que se observa a lo largo del Africa central, puesto que las áreas quemadas aparecen ennegrecidas en pasadas subsiguientes. La superficie total de tierra limpiada de este modo ha sido medida a partir de estas imágenes del Landsat.

Al parecer, las únicas regiones donde abundan los incendios de origen agrícola corresponden a los países menos desarrollados. Presumiblemente esta práctica está relacionada con la falta de maquinaria adecuada para preparar la tierra y proceder a nuevas plantaciones; en consecuencia, se utiliza el fuego como ayuda para limpiar los restos de las cosechas de la estación precedente. Desde luego, algunos de estos restos requieren la utilización del

DISTRIBUCION DE LA POBLACION de los Estados Unidos. Se observa en el montaje de tres imágenes nocturnas que se reproducen arriba. A efectos de comparación en la parte inferior se reproduce el mapa de población, obtenido con un ordenador a partir de los datos suministrados por la Oficina del Censo.

fuego incluso en circunstancias tales en que se podrían utilizar otros tratamientos como alternativa. Sin embargo, resulta un tanto sorprendente comprobar la importancia que tiene la utilización del fuego en aquellos países en los que se carece de maquinaria agrícola moderna. Hace un siglo se recurría a fuegos similares a éstos para ayudar a aclarar los enormes bosques que cubrían las regiones centrales de los Estados Unidos.

Los incendios naturales se ponen de manifiesto también en las imágenes que proporcionan los satélites de las Fuerzas Aéreas, por ejemplo, en el noroeste de Australia *(véase la ilustración de la página siguiente).* Al contrario de lo que sucede con los incendios de las tierras cultivadas, la extensión de estos incendios es muy grande. Hay dos aspectos de tales incendios que ratifican su carácter accidental: generalmente aparecen en regiones remotas no cultivadas y son tan grandes que probablemente serían difícilmente controlables por el hombre, si es que pudieran serlo en alguna medida. (Por otra parte, se ha indicado que los aborígenes de dicha región quemaban el monte bajo para mantener las tierras de pasto de los canguros; es posible que esta práctica se siga todavía aplicando a gran escala en la actualidad.) Al parecer, estos grandes incendios destruyen la vida salvaje; personalmente he recibido una consulta del Servicio de Vida Salvaje de Australia, interesándose por la utilización de estas imágenes de satélites en el control de los incendios periódicos de esta región. Evidentemente, este proyecto no sólo es factible sino incluso deseable como forma de evaluar el daño producido por los grandes incendios naturales. Excepto en épocas de alta nubosidad, la evolución de los fuegos se podría detectar diariamente.

Si otra civilización observara la Tierra desde el espacio no llegaría a la conclusión de que aquí nos enfrentamos a una escasez de energía; en realidad, la visión nocturna de la Tierra pone de manifiesto un gran despilfarro, que indica una superabundancia de energía. El derroche se debe en parte a lo reacios que se muestran los países desarrollados (en particular, precisamente aquellos que están bien iluminados por la noche) a depender de las importaciones de energía debido a la amenaza que para su seguridad nacional y bienestar económico supone dicha dependencia.

Las luces de las ciudades del mundo desarrollado y las llamas de gas del mundo menos desarrollado son un hecho relativamente reciente en la historia de la humanidad. Hace tan sólo un siglo, las imágenes nocturnas de la Tierra observada desde el espacio no hubieran puesto de manifiesto ni ciudades ni llamas de gas y



LA BRILLANTE AURORA que se extiende sobre todo el Canadá y tan al sur que llega hasta la región de los Grandes Lagos contituye el hecho más destacable de esta fotografía, tomada por el satélite meteorológico de las Fuerzas Aéreas norteamericanas durante la noche del 18 de abril de 1974. Las luces de las ciudades de los Estados Unidos se distribuyen a lo largo del centro de la imagen. Las pequeñas luces que se observan al oeste de México, en el lado inferior izquierda de la figura, están originadas por los fuegos agrícolas. La gran llama de gas de la parte inferior derecha está asociada con el campo petrolífero Reforma, en México.

mucho menos incendios agrícolas. ¿Qué aspecto ofrecerá la cara obscura de la Tierra dentro de otros 100 años? Yo me atrevería a predecir que todas las regiones habitables tendrán el mismo aspecto que los Estados Unidos en la actualidad, es decir, estarán cubiertas por las luces de las ciudades. Por el contrario, no se detectara, la presencia de las llamas de gas. El gas y el petróleo se habrán convertido probablemente en materias primas químicas y sólo se utilizarán como combustibles en ciertos casos específicos.

¿Por qué razón las compañías explotadoras de los campos petrolíferos no detienen el derroche que supone la combustión del gas instalando un sistema de licuefacción? La respuesta a esta pregunta tan inmediata presenta aparentemente tres aspectos diferentes. En primer lugar, habría que señalar que, de hecho, se están invirtiendo miles de millones de dólares en equipos de licuefacción y en los barcos necesarios para transportar el gas líquido hasta sus lugares de consumo. La importancia de este esfuerzo ha sido dificultada por las declaraciones de los líderes de los países más desarrollados poniendo de manifiesto su aversión a la compra de dicho gas ante el temor de llegar a ser demasiado dependientes de él. La balanza de pagos y la preocupación por la seguridad nacional constituyen otras tantas causas indirectas de la combustión del gas residual.

Un segundo factor desempeña también un papel importante en toda esta situación. Los sistemas de licuefacción actuales resultan económicamente rentables únicamente en el caso de que la correspondiente planta sea grande. Los pequeños centros de combustión tienen que conectarse a los grandes centros; de lo contrario, el gas no se puede recuperar de forma económica. Se hace necesario proyectar una planta de licuefacción gaseosa de pequeño tamaño que sea rentable. Quizá las imágenes de los satélites pongan de manifiesto que las pequeñas llamas son tan abundantes que justifican una inversión sustancial en el desarrollo de dichas plantas.

El tercer factor a considerar en la resolución del problema que estamos analizando consiste sencillamente en el hecho de que en muchas de estas llamas se elimina un gas altamente corrosivo, para el que no se ha encontrado todavía un uso adecuado.

En relación con los fuegos agrícolas, cabe señalar la posibilidad de que este tipo de combustiones a gran escala produzcan la extracción de ciertos elementos nutritivos del suelo, dispersándolos en la atmósfera. Aunque se reconozca que el fuego es necesario para la supervivencia de algunos sistemas ecológicos, el hecho de que las regiones cuya productividad agrícola es más baja sean también aquellas en donde la práctica de los fuegos agrícolas se aplique con mayor frecuencia parece una coincidencia demasiado grande. Quizás el uso del fuego en la preparación de la tierra para la plantación tenía más sentido en los tiempos en que la población no era tan elevada. Tal como indican claramente las imágenes de los satélites, las alternativas a la combustión de los campos que conserven los elementos nutritivos del suelo podrían aplicarse en todo el mundo.

FUEGOS DE ORIGEN NATURAL. Parecen ser los responsables de las grandes manchas brillantes que se observan en esta imagen, tomada durante la noche por un satélite que sobrevolaba Australia. Las dimensiones de estos incendios son muchos mayores que las de los fuegos controlados que se observan en las regiones cultivadas. Las rayas brillantes y horizontales que se destacan sobre las nubes próximas están producidas por relámpagos. El extremo sudoccidental de Australia aparece claramente iluminado por la luz de la Luna.

Aprendizaje y memoria de las abejas

La abeja puede aprender en seguida y recordar por largo tiempo el color y el olor de las flores que le proporcionan néctar o polen. Se está desvelando la base nerviosa de tal comportamiento

Randolf Menzel y Jochen Erber

Si nos fijamos en una abeja determinada que va de flor en flor observaremos que siempre escoge el mismo tipo de flor y no se preocupa de las inmediatas que posean otros colores, formas o aromas, que atraerán a otras abejas. El insecto no podría actuar de esta manera si no fuera capaz de dos cosas: distinguir una flor de otra y aprender qué flores pueden ofrecerle néctar o polen y, por tanto, merece la pena visitar. Dicho de otro modo, la abeja posee memoria y facultad de aprender.

Con nuestros colegas de la Universidad Libre de Berlín hemos venido investigando estos recursos de la abeja con la esperanza de poder descubrir así la base celular del aprendizaje y la memoria. Nuestros estudios son a la vez de comportamiento (nos dedicamos a su observación en varias situaciones experimentales para determinar hasta dónde llega su repertorio etológico) y fisiológicos (analizamos la anatomía y funcionamiento del cerebro de la abeja al objeto de descubrir qué partes del mismo intervienen en el aprendizaje y la memoria). Y nos encontramos con que, aunque la abeja sea una máquina de comportamiento detalladamente programado, sus mecanismos de aprendizaje y memoria son bastante parecidos a los que operan en organismos mucho más evolucionados, hombre incluido.

Aunque para muchos y atentos observadores del pasado (las notas y testimonios escritos van desde Aristóteles hasta Charles Darwin) era ya obvio que las abejas podían ver los colores, oler el aroma de las flores y recordar qué tipo de flor habían visitado la vez anterior, la prueba experimental tuvo que esperar hasta 1910, cuando Karl von Frisch, de la Universidad de Munich, llevó a cabo una prueba sencilla, que sigue siendo uno de los pocos experimentos realmente cruciales y decisivos en el estudio del comportamiento animal. El científico atrajo abejas recolectoras hacia una mesa, ofreciéndoles una solución de miel. Cuando había suficientes abejas que visitaban el lugar con regularidad, el plato que contenía la miel se colocaba sobre un trozo de cartulina azul, con lo que las abejas veían el color azul mientras se acercaban a la mesa y chupaban la miel.

En ese momento, von Frisch lo tenía ya todo dispuesto para su prueba. Rodeó la cartulina azul con distintos trozos de cartulina del mismo tamaño, pero cuyo color variaba desde el blanco hasta el negro, pasando por distintas intensidades de gris. Encima de cada trozo de cartulina, incluida la azul, había un plato vacío. Las abejas, que ya habían sido acostumbradas a que podrían encontrar comida en la mesa, continuaban mirando y casi todas ellas limitaban la búsqueda a la cartulina azul.

El experimento apoyaba vigorosamente los dos postulados que se han mencionado más arriba. Las abejas habían distinguido el color azul de entre todos los demás colores, lo que indicaba que podían ver el azul como color y también que habían aprendido a tomar el color (en tales circunstancias) como señal de comida. Para confirmar el segundo descubrimiento era necesario demostrar que las abejas no habían sido atraídas espontáneamente por el color. Von Frisch consiguió esto a través de varios sistemas, al igual que lo hicieron otros después de él. Efectivamente, su experimento se ha venido repitiendo con diversos colores, olores y dibujos en blanco y negro.

Por consiguiente, se conocen bien los distintos aspectos de la actividad de la abeja. Mas, por sorprendente que pudiera parecer, el proceso de aprendizaje que interviene en todos estos experimentos no se estudió hasta hace unos 10 años. Veremos que dicho proceso posee varias facetas interesantes que lo hacen atractivo de abordar, enfocado desde el prisma de un análisis neuronal de la formación de la memoria en el cerebro de la abeja.

Vaya la primera pregunta. ¿Con qué precisión seleccionan el tipo de flores, néctar y polen que pueden recoger? Esta pregunta se puede contestar con rapidez y justeza sin necesidad de ir siguiendo a una abeja en su vuelo de flor en flor. Cuando un individuo practica uno de sus periódicos retornos a la colmena, lleva un recuerdo de lo que ha escogido en los dos montones de polen (uno en cada una de las patas posteriores). Observando el contenido de la masa de polen de cada abeja se descubre que el polen recogido por los individuos en una sola salida es total o casi totalmente de un mismo tipo de flor. Puesto que una abeja llega a visitar hasta 500 flores por excursión, el grado de precisión con el que escoge resulta, pues, elevado.

Esta precisión en la respuesta, unida a la rapidez con que las abejas pueden acostumbrarse a reaccionar ante diversos estímulos, convierten a nuestros insectos en especialmente idóneos para realizar los experimentos que nos proponemos. También resulta de considerable ayuda el hecho de que todas las abejas obreras de una colmena son hermanas y, por tanto, genéticamente muy parecidas; sin olvidar que están motivadas para la búsqueda de alimento, de suerte que cada abeja toma un gran número de decisiones (de 1500 a 2000 por día).

Al objeto de abordar de manera directa los sucesos neuronales que subyacen al aprendizaje y la memoria, examinamos en primer lugar, y muy detenidamente, el proceso de aprendizaje siguiendo una metodología etológica. A este respecto, insistimos en aquellos experimentos, de acuerdo con los cuales, las abejas se sometían al entrenamiento de responder a diferentes colores. (Otros investigadores han empleado métodos diferentes. Así, Martin Lindauer y sus colegas de la Universidad de Würzburg ensayaron con abejas sometidas a pruebas de olores o, también, a horario diario preestablecido; Rüdiger Wehner, de la Universidad de Zurich, las entrenó para distinguir entre diferentes dibujos en blanco y negro.)

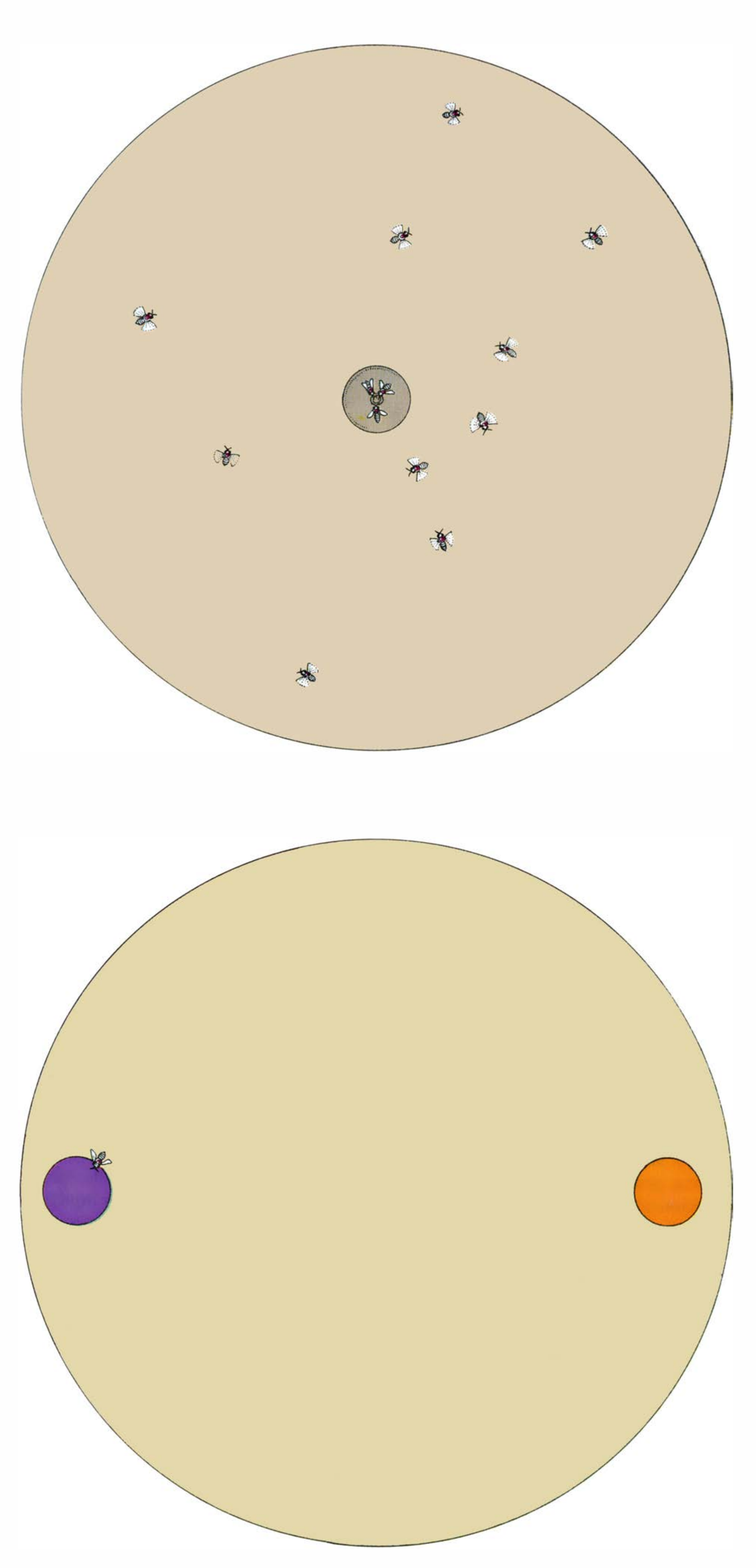
Nuestro sistema para practicar el entrenamiento con los colores comprendía un aparato en el que se podía utilizar luz monocromática de intensidad regulable para iluminar los discos de cristal fijos. Las

abejas, que estaban marcadas individualmente, comían en un disco de cristal no iluminado y situado en el centro de la mesa de experimentos de color gris. Por medio de una "danza zumbona", descubierta por von Frisch, las abejas informaban a sus compañeras de colmena de la fuente de alimento encontrada. A una nueva que llegara en respuesta a esta información se le permitía visitar, por tres veces, un disco no iluminado, de manera que podía familiarizarse con el dispositivo. Después de tres visitas que le proporcionaban comida, la abeja estaba motivada para buscar alimento aun cuando no se le ofreciera más. Mientras una nueva abeja efectuaba sus tres primeras visitas, todas las demás abejas se mantenían fuera del aparato, guardadas en una jaula; así pues, unicamente la abeja motivo del experimento estaba volando entre la colmena y el aparato.

Este primer ensayo evalúa la preferencia de la abeja por la elección espontánea. Por primera vez, la abeja objeto del experimento se encuentra con dos discos de cristal de colores diferentes. Ninguno de los cuales contiene alimento, y ambos están equidistantes del centro (lugar primitivo de alimentación). La abeja se acerca a los discos, aterriza sobre ellos y busca comida. Si la frecuencia de elección no es igual para ambos colores, la intensidad de un color u otro se ajusta, hasta que el insecto no denota ninguna preferencia por ninguno de los dos discos.

Después de un período de prueba de escasos minutos se la recompensa con comida en uno de los dos colores. Cuando vuelve, se hace otra prueba como la del juicio de preferencia espontánea. Se la gratifica de nuevo, y así sucesivamente. Con este procedimiento obtenemos curvas de aprendizaje de abejas tomadas una a una; por tanto, nos permiten evaluar cuantitativamente los factores que intervienen en el comportamiento de aprendizaje de las abejas.

Bastan pocas pruebas para que aprendan los colores. El violeta (a longitudes de onda de 400 a 420 nanómetros) es el co-



ENTRENAMIENTO para el color por parte de las abejas. Para llevarlo a cabo, los autores se sirvieron de un aparato que les permitía iluminar los discos de cristal (de siete centímetros de diámetro) con luz monocromática de intensidad regulable. En primer lugar se alimenta a las abejas marcadas una a una (*arriba*) en un disco no alumbrado que se halla en el centro de la mesa de entrenamiento, de color gris. Con su "danza zumbona" las abejas informan a sus compañeras de colmena de la situación de la fuente alimentaria. A una abeja que responda se le da de comer en el disco central; se observa cuál de los dos colores que equidistan del disco central escoge espontáneamente (*abajo*); luego se le recompensa con comida en el disco central, iluminado ya por uno de los dos colores. Por pequeña que sea, la recompensa hace que el insecto vuelva siempre hacia ese color.



APRENDIZAJE DEL COLOR por parte de las abejas. Estas aprenden algunos colores con bastante rapidez, pero otros les cuestan más, aunque pueden aprender todos los que les resulten visibles. En esta composición, se representa la flor del rosal silvestre en el centro, tal como la ve el ojo humano; a continuación, y en sentido horario comenzando arriba, aparece el orden de colores que la abeja asocia mejor con una recompensa de comida. Así, el color que se aprende antes es el violeta intenso, luego el azul, el ultravioleta (*que se muestra en gris*), el verde y el anaranjado. El color verdeazulado, que constituye el color de fondo natural para las abejas, es el que más les cuesta aprender. (Las ilustraciones son de P. J. Wynne.)

lor que recuerdan con mayor rapidez; una sola recompensa es suficiente para que se produzca una alta frecuencia de elecciones correctas. El verde-azulado (490 nanómetros) se aprende con mayor lentitud. Las diferencias características de las curvas de aprendizaje no dependen de factores tales como la intensidad de los colores (siempre que la intensidad alcance por lo menos la mitad de una unidad logarítmica por encima del umbral de percepción de la abeja), la longitud de onda del color que se coloca como alternativo, el tipo de preentrenamiento, la edad de la abeja objeto del ensayo, la época (período del día o bien la estación) y el tiempo atmosférico.

Podría esperarse que la dependencia del aprendizaje respecto de la longitud de onda mostrara una estrecha relación con el sistema de visión de los colores. No sucede así. El sistema de visión de los colores de la abeja es tricromático, y está dotado de receptores que poseen su sensibilidad máxima a la radiación ultravioleta (350 nanómetros), la luz azul (440 nanómetros) y la luz verde (540 nanómetros). Como ya demostrara, hace veinte años, Karl Daumer, de la Universidad de Munich, el ultravioleta resulta ser el color más vivo y conspicuo para las avejas. Y según cabría esperar de tal sistema, la capacidad de distinguir entre los colores es mayor para el violeta y el verde azulado, toda vez que la sensibilidad de los receptores se halla más acusada en esa zona y allí coinciden en punto a sensibilidad. Efectivamente, Otto von Helversen, de la Universidad de Friburgo, demostró que dicha hipótesis era correcta. Tales sensibilidades son precisas para una cuidadosa identificación de las flores en las que predomina el violeta y el azul, siendo éstas las especializadas en atraer a los insectos para realizar la fecundación cruzada.

A pesar de que las abejas son muy sensibles a la radiación ultravioleta, no es el color lo que mejor aprenden. Para la abeja, la radiación ultravioleta representa la radiación celeste, y no parece que deba tomarse como un puro color indicativo de suministro alimenticio. Se aprende mucho antes si se mezcla con luz de longitud de onda larga para formar el "púrpura de la abeja". El verde azulado representa el color de fondo, y no se le entiende como indicador de comida. La interacción evolutiva entre las flores y los insectos que las polinizan no sólo ha resultado en una adaptación de los colores de las flores y la percepción de estos colores por parte de las abejas, sino también en ciertas disposiciones prefijadas de aprendizaje en estos individuos. Interpretación que se apoya en los trabajos de Lindauer sobre el aprendizaje de los olores por parte de las abejas; obviamente, los olores comunes se aprenden antes que los olores artificiales o los que son poco frecuentes.

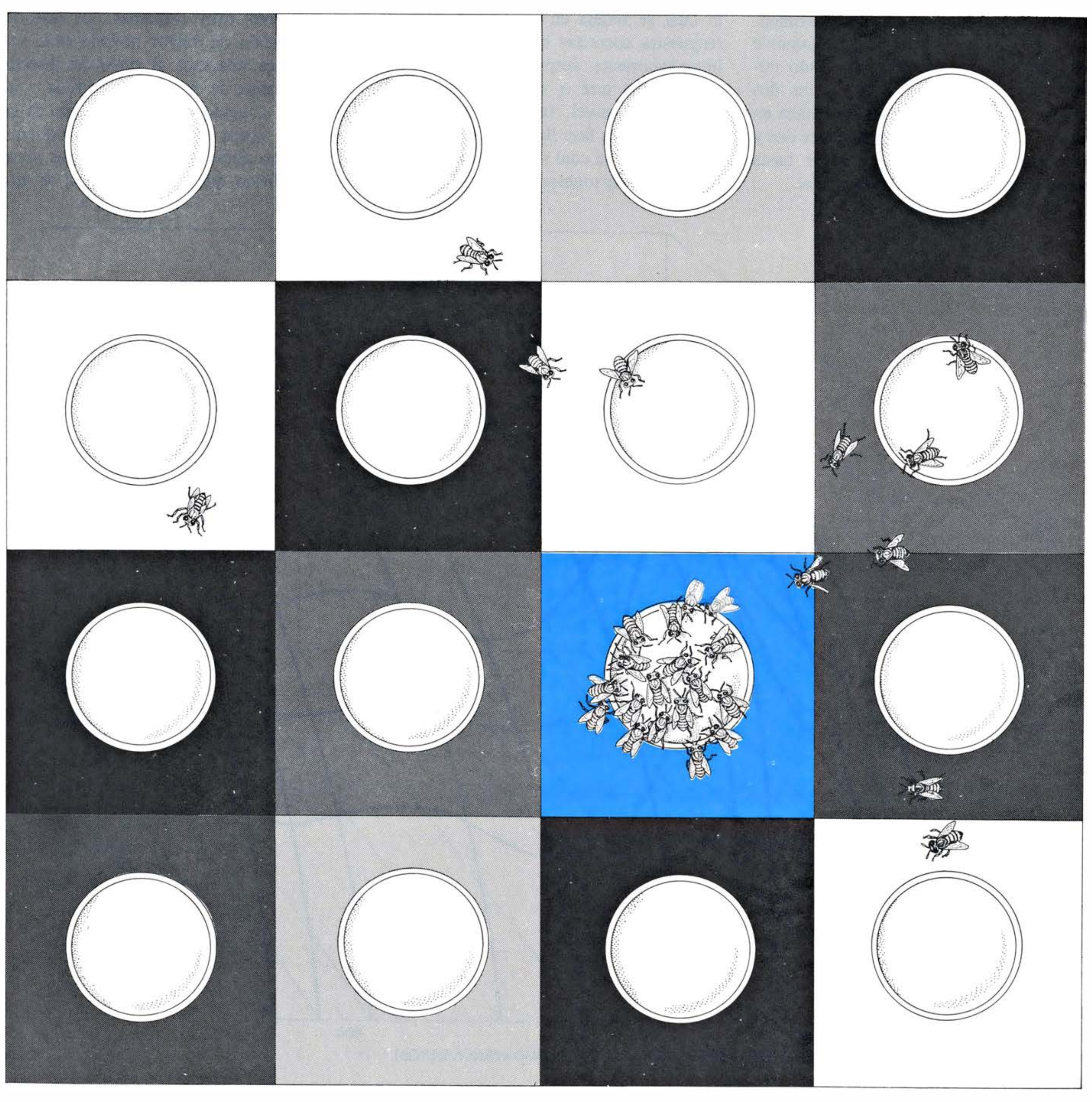
En este contexto debe destacarse la reciente tendencia ente los psicólogos experimentales a realzar el valor de la perspectiva etológica; con otras palabras: el comportamiento se comprende mejor en el marco del entorno y evolución naturales del animal. El aprendizaje del color y el olor por parte de las abejas aparecen como muestras de comportamiento, de cuyo examen se saca el máximo rendimiento en este enfoque. Como etólogo que era, von Frisch interpretó de una manera intuitiva los resultados que obtenía mirándolos a través de este prisma, hace ya 40 años. Hoy podemos separar cuantitativamente la configuración del sistema perceptivo de la abeja (el programa de comportamiento para recabar información de un contexto específico) de los programas nerviosos para procesar dicha información en la memoria almacenada.

¿Qué estabilidad posee la memoria de las abejas respecto de las señales de alimentación? Referencias aleatorias de apicultores hablan de una extraordinaria estabilidad y no pequeña durabilidad. A Lindauer debemos una de las primeras observaciones cuantitativas fechada en 1963, cuando descubrió que las abejas llegaban espontáneamente a los centros de alimentación tras cinco meses de ausencia durante el invierno. Al objeto de conseguir una valoración cuantitativa más refinada, sometimos a las abejas a un entrenamiento para el color azul o el amarillo, dándoles de una a tres recompensas; luego se encerraban en jaulas de malla individuales dentro de la colmena por periodos de distinta duración y se probaba entonces su comportamiento selectivo. Despueś de haberles dado una recompensa, la frecuencia de una elección determinada seguía manteniéndose elevada a lo largo de varias horas, para bajar al día siguiente y alcanzar el nivel de elección espontánea a los cinco o siete días. Sin embargo, después de tres recompensas, no advertimos ninguna reducción en el número de elecciones correctas en un periodo de 13 días, aproximadamente, que son los que una abeja suele sobrevivir enjaulada. Parece claro que las abejas sólo necesiten de unas pocas recompensas para poseer una memoria dotada de un alto grado de estabilidad.

El cuadro cambia cuando se permite a

las abejas comenzar de nuevo el aprendizaje. Pasarán a otra fuente de alimentación marcada con un color distinto, si bien la velocidad del cambio depende del número de recompensas recibidas en el color ya no premiado. En esta situación se puede observar un curioso comportamiento bien conocido a partir de los estudios con vertebrados. El animal verifica el cambio más deprisa si ha tenido un largo periodo de entrenamiento (25 o más recompensas) para el color inicial. Otro hallazgo da cuenta de que, cuando hay dos signos que comportan recompensas a distinta tasa, el animal optará por uno u otro según la probabilidad de recompensa. (Este efecto ha sido minuciosamente estudiado por Bernhard Schnetter, de la Universidad de Würzburg). Otro caso a analizar es el entrenamiento simultáneo para más de un color. Las abejas pueden aprender dos colores a la vez y pueden distinguirlos de entre los demás. Pero si se recompensan más de dos colores simultáneamente, el individuo no puede diferenciar ya entre ellos. Todo esto nos revela algo sobre la lógica del diseño de sus programas de aprendizaje.

Esos efectos de aprendizaje, complejos y duración, muestran interés desde el punto de vista de la etoloe la etología y la psicología comparativa, pero no son los adecuados para un análisis de los procesos neuronales que subyacen al comportamiento. Así fue como centramos nuestros trabajos en el estudio de los procesos neuronales que intervienen durante y después de cada asociación de una señal de color, u olor, y una recompensa. Decisión ésta que resultó ser afortunada porque inmediatamente comenzaron a abrirse nuevas perspectivas sobre el proceso de aprendizaje.



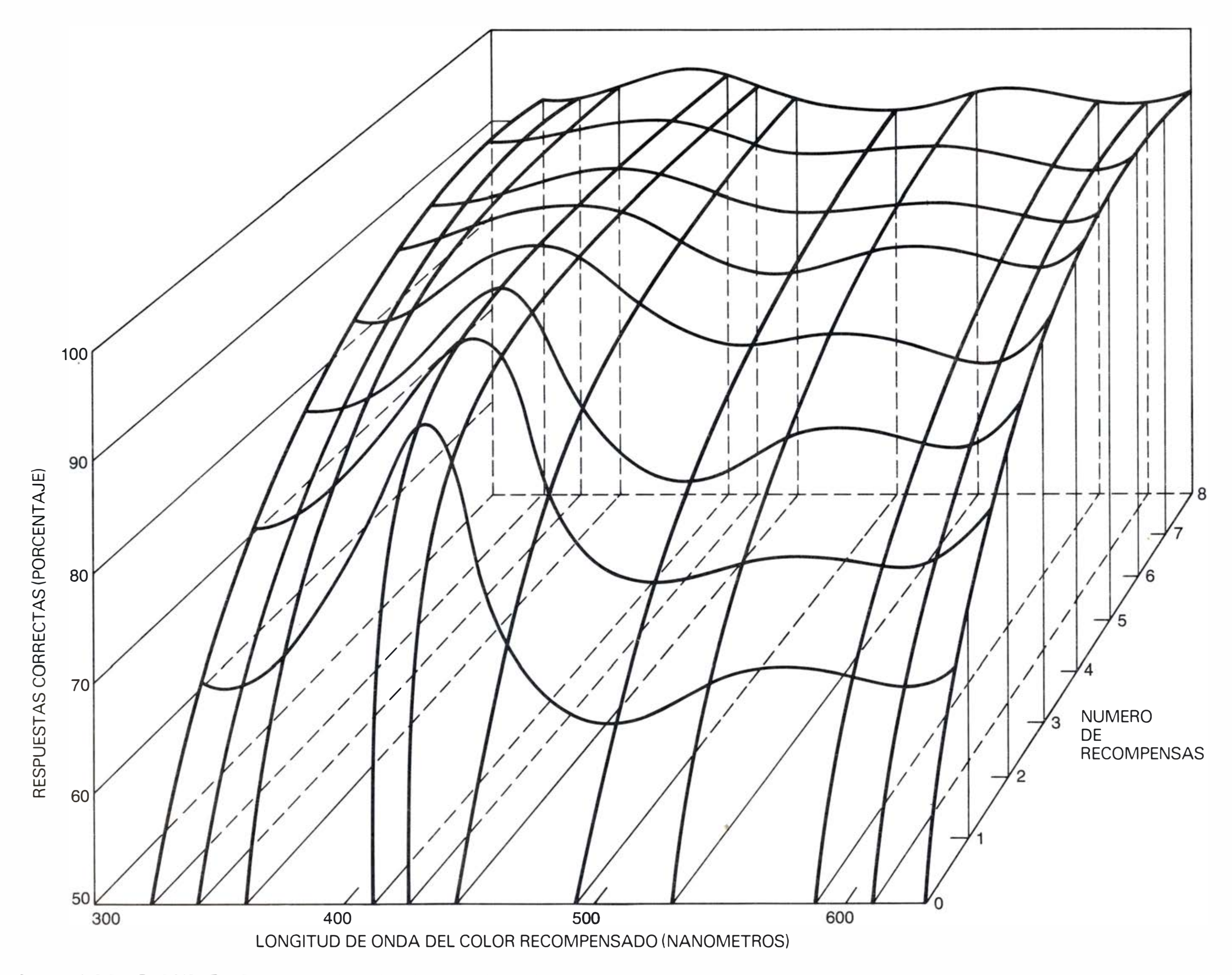
LA PRUEBA DEL COLOR realizada por Karl von Frisch en 1910 estableció la capacidad de las abejas para asociar los colores con el alimento y recordar el color así aprendido. Colocó un plato que contenía agua con miel en una mesa. Las abejas se sentían atraídas a él. Después puso el plato sobre un trozo de cartulina azul; por tanto, el azul era el color que veían las abejas mientras se acercaban al plato y mientras sorbían la miel. A continuación von Frisch rodeó la cartulina azul con otras cartulinas del mismo tamaño pero de color blanco, negro o de distintas tonalidades de gris. Cada cartulina, la azul incluida, tenía un plato vacío encima. Se las había acostumbrado a que hubiera comida en la mesa y por ello seguían viniendo. La mayoría sólo miraba la cartulina azul.

Conseguimos fijar un primer resultado especialmente provechoso: tras cada recompensa única, el nivel de respuesta es independiente de la cantidad que se dé como recompensa, siempre que se tengan en cuenta ciertas limitaciones umbrales. Hasta una recompensa que dure sólo 100 milisegundos, y que tenga una baja concentración de sucrosa, puede bastar para que se produzca un cambio significativo en la respuesta.

Obtuvimos otro resultado de interés al entrenar a las abejas dándoles una recompensa única de corta duración (dos, cinco o diez segundos). Las abejas sometidas a prueba mostraron inmediatamente después de la recompensa un eleado número de elecciones correctas. A los dos minutos, sin embargo, el índice bajaba en picado, alcanzando un mínimo entre dos y tres minutos, para volver a subir hasta situarse de nuevo en el nivel inicial.

Este tipo de relación entre la formación de la memoria y el tiempo se encuentra en otros animales y en el propio hombre. Los mecanismos fisiológicos que subyacen a la asociación entre una señal y una recompensa (o un castigo) necesitan tiempo (segundos, minutos e incluso horas), y lo más seguro es que atraviesen sucesivas fases en las que el substrato neuronal de la memoria sea diferente. En el caso típico, el primer almacenamiento se da en la memoria sensorial, de donde rápidamente puede desaparecer si no se la refuerza. Pero si ocurre lo contrario, la información se registra en una memoria a corto plazo, lo cual se refleja en el elevado nivel de respuestas correctas que opera una abeja inmediatamente después de haberla recompensado una o varias veces. El ascenso en el nivel, tras la previa caída, representa una fase llamada de consolidación, durante la cual se transfiere la información a una memoria de largo plazo.

Tal como cabía esperar, las fases de aprendizaje de la abeja se hallan perfectamente adaptadas a las condiciones naturales de aprendizaje. Muchas flores ofrecen sólo cantidades pequeñas de néctar, de suerte que la abeja se entretiene chupando breves segundos y pasará a la siguiente flor en pocos segundos también. Pero si las flores se encuentran muy dispersas, el animal puede tardar hasta un minuto en dar con otra flor parecida. La alta tasa inicial de elecciones correctas revela un mecanismo de corto plazo que le asegura a la abeja la capacidad de hallar otra flor similar incluso después de haber recibido recompensas muy pequeñas. La fase de consolidación de mayor periodo está, según parece, adaptada al vuelo del insecto entre la zona de captura de comida y la colmena. Cuando la abeja torna desde la colmena, opera de nuevo un elevado índice de elecciones correctas; ello nos pone de manifiesto que la información se en-



CURVAS DE APRENDIZAJE, que simbolizan la velocidad a la que las abejas aprenden los distintos colores. Se han construido a partir del promedio entre muchos individuos. El promedio de respuestas correctas sube con el número de recompensas alimentarias. En estos experimentos el color que se daba como alternativa al color premiado era el color complementario: para el ultravioleta de 361 nanómetros era el verdeazulado de 494; para el violeta de 428 nanómetros el anaranjado de 590 y, para el azul, era el "púrpura de las abejas", que es una mezcla de 80 por ciento naranja y 20 por ciento ultravioleta (*última curva a la derecha* **). El aprendizaje más raudo se da con el violeta a 408 y 418 nanómetros; el más lento se da en el verdeazulado a 494 nm. Si se entrenan para el violeta, las abejas alcanzan un nivel de precisión del 90 por ciento después de una recompensa; para el verdeazulado precisan cinco recompensas.**

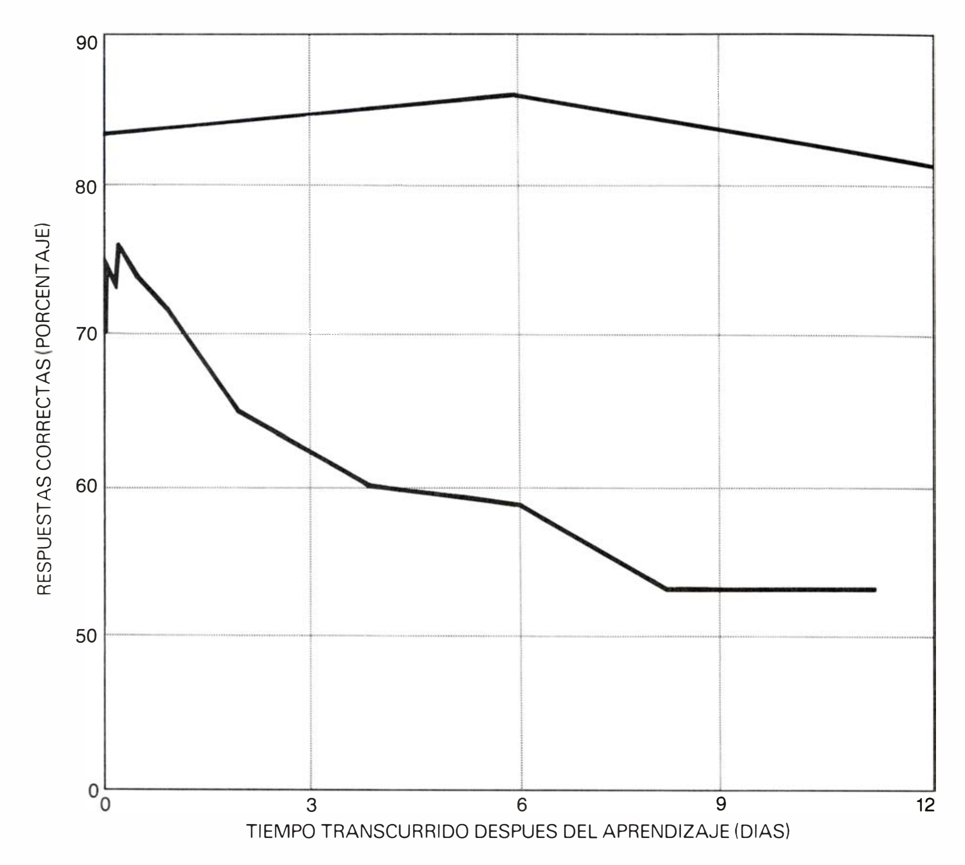
cuentra en una memoria o registro de larga duración.

¿Cuánto dura almacenada una señal de color en la memoria sensorial de la abeja? La clave de la respuesta se halla en el establecimiento de las condiciones bajo las cuales el color seguirá asociado a la comida, aunque aquél ya no se vea. En los experimentos que hemos acometido en esta línea de trabajo, hemos observado que la señal de color percibida por la abeja en el intervalo de los dos segundos antes de recompensarla se aprende con la misma fuerza que si la señal cromática hubiera estado presente a lo largo de todo el periodo de aproximación y succión del néctar. Si se dispone la señal de color una vez que la abeja ha empezado ya a chupar, no se la asociará a la comida.

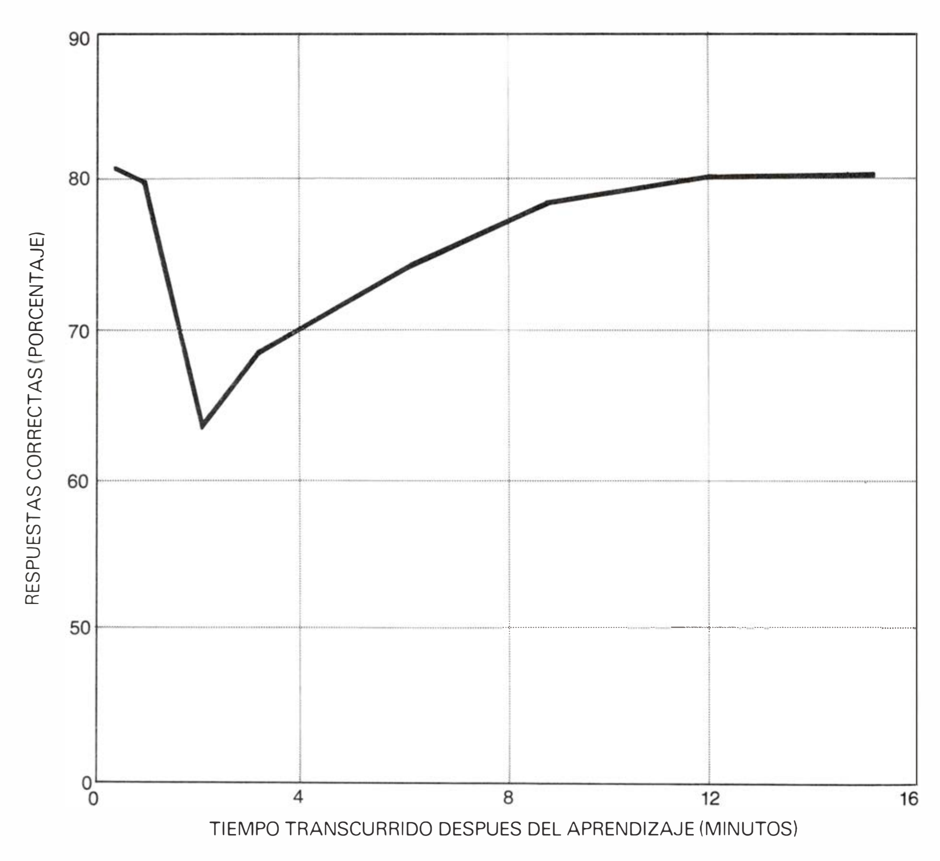
Elisabeth Opfinger observó en 1931, estudiando a la sazón con von Frisch, que las abejas aprenden un color sólo si lo ven mientras se van acercando a la flor; los colores contemplados mientras están comiendo o a la vuelta no se fijan. Ahora sabemos ya que una señal de color debe hallarse entre los tres segundos que preceden a una libación y los 0,5 segundos de haber comenzado la misma para que pueda recordarse. Puesto que basta una sola libación de dos segundos para que la abeja muestre un alto nivel de aprendizaje, el periodo total de aprendizaje inicial (memoria sensorial y periodo de asociación más la duración de la recompensa) resulta estar comprendido entre cuatro y cinco segundos. Estamos pues ante una situación adecuada para explorar las fases temporales de almacenamiento de memoria, que se iniciaron con una sola recompensa.

¿Acaso las fases de corto y de largo plazo reflejan dos mecanismos de almacenamiento fisiológicamente distintos? Hemos sometido a las abejas a varios procedimientos que se sabe bloquean la actividad neuronal del cerebro. Todos ellos impiden la formación de memoria de corta duración o a corto plazo. (Una vez se ha recuperado, las abejas así tratadas se comportan exactamente igual que los individuos no sujetos a experimentación. Es más, las pruebas de control excluyeron la posibilidad de que los métodos experimentales seguidos causaran un refuerzo negativo.)

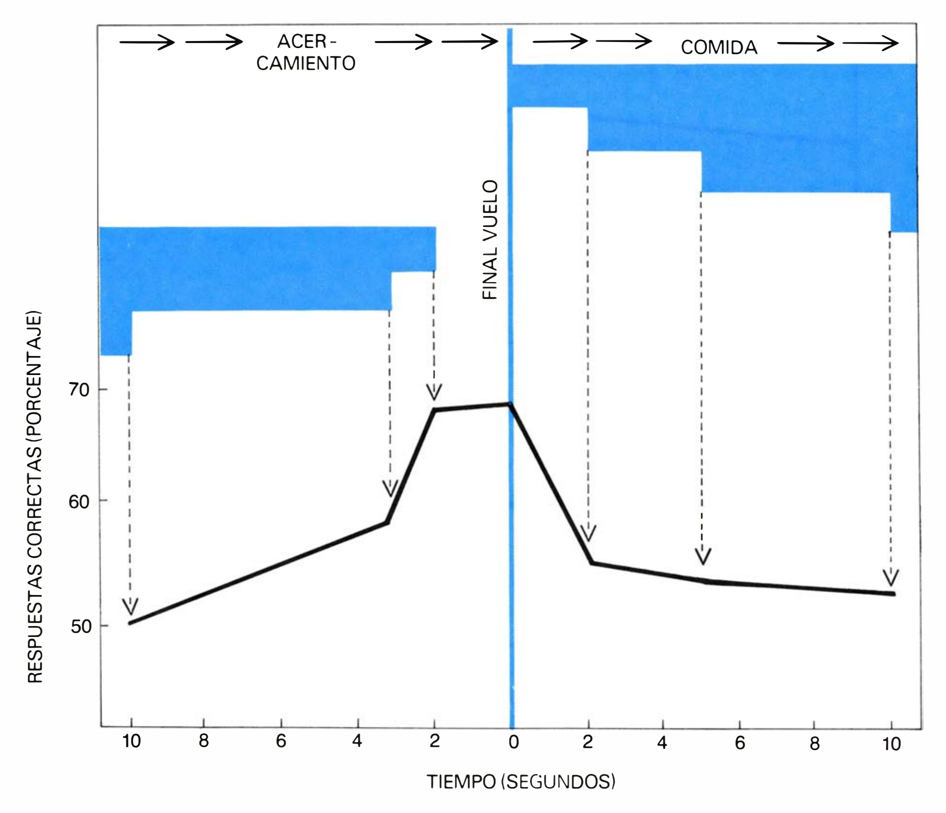
Hemos observado la existencia de correlación entre el periodo inicial de caída o pérdida de memoria y el periodo de sensibilidad a los procedimientos de bloqueo. Los resultados indican que durante un periodo inicial de pocos minutos el almacenamiento de memoria se funda en unos mecanismos fisiológicos que son distintos de los implicados en un almacenamiento posterior. Casi se podría tener la seguri-



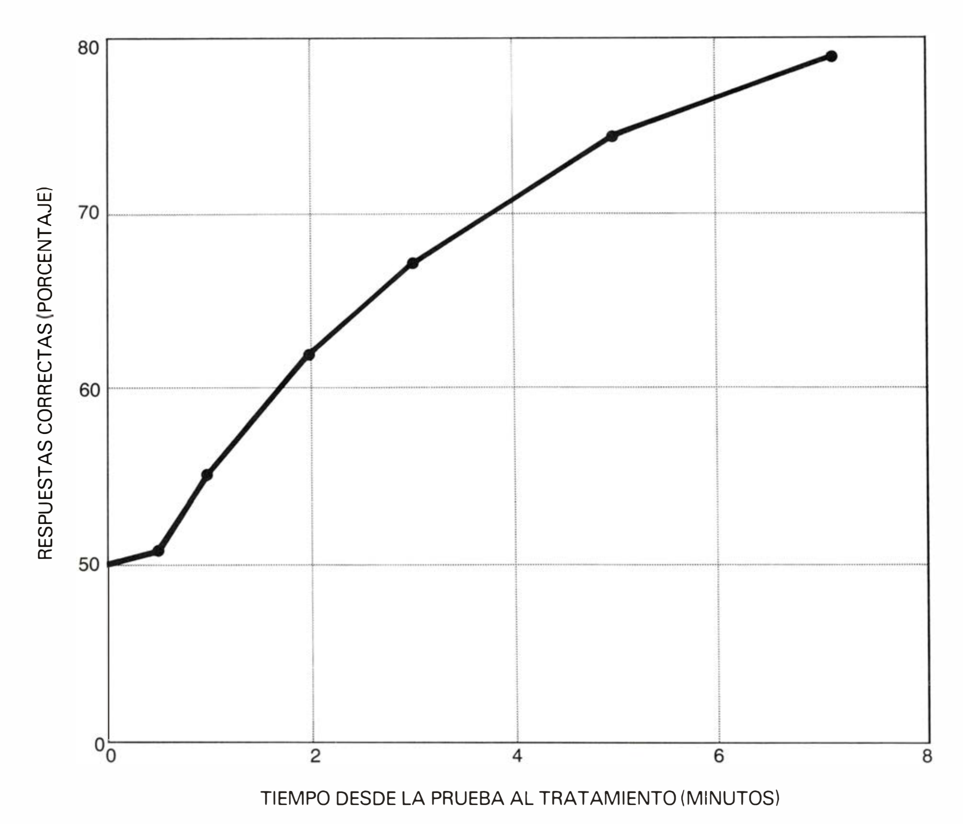
GRAFICAS DE LA ESTABILIDAD DE MEMORIA en las abejas. Si recompensamos a éstas, por tres veces, a propósito de un color determinado, y se las mantiene luego en la colmena durante un tiempo que puede llegar hasta las dos semanas, para evitarles nuevos aprendizajes, conservarán un elevado índice de respuestas en adelante (*arriba*). Con una sola recompensa, el nivel de respuestas aumenta en un principio, durante un periodo que se llama fase de consolidación. Al día siguiente, las respuestas comienzan a descender; se llega al nivel de opciones espontáneas a los seis días. El disco recompensado era el azul.



COMPORTAMIENTO SELECTIVO de las abejas subsiguiente a un periodo de prueba para el aprendizaje. Las abejas sometidas a prueba inmediatamente después del aprendizaje eligen el color premiado con fina precisión. Dos minutos más tarde, el número de respuestas adecuadas decae, para volver a aumentar.



PERIODO DE ASOCIACION entre un color y la recompensa alimentaria. Dura muy poco tiempo y queda claramente dibujado. Un color (en este caso se trataba del azul) que era visto por la abeja durante un tiempo superior a los cuatro segundos antes de posarse o superior a los dos segundos después de haberse posado y empezado a comer no va asociado a la recompensa. Una señal de color, suprimida dos segundos antes de que empiece la recompensa sí resulta aprendido. Los experimentos manifiestan la existencia de una memoria sensorial que pervive hasta tres segundos. La información recabada debe reforzarse con alimento si tiene que procesarse en la memoria de corto y largo plazo, o almacenamiento de corta y larga duración.



CONDUCTA DISRUPTIVA, que produce los efectos señalados en la formación precoz de memoria en las abejas. Después de haberse determinado una elección espontánea de un color por parte de la abeja, se la recompensó bajo una luz azul, durante 30 segundos. Se le congeló luego el cerebro inmediatamente o después de distintos intervalos señalados por círculos. La curva indica el porcentaje medio de respuestas correctas una vez que las abejas se hubieran recuperado del tratamiento a que habían sido sometidas. La destreza de las abejas que habían padecido la congelación era la normal al cabo de siete minutos. La congelacion anterior interfería evidentemente con la formación de la memoria de corto plazo (o duración).

dad de que los mecanismos fisiológicos precoces están constituidos por actividades ordenadas de las neuronas, ya que nuestro ensayo bloqueaba dichas actividades.

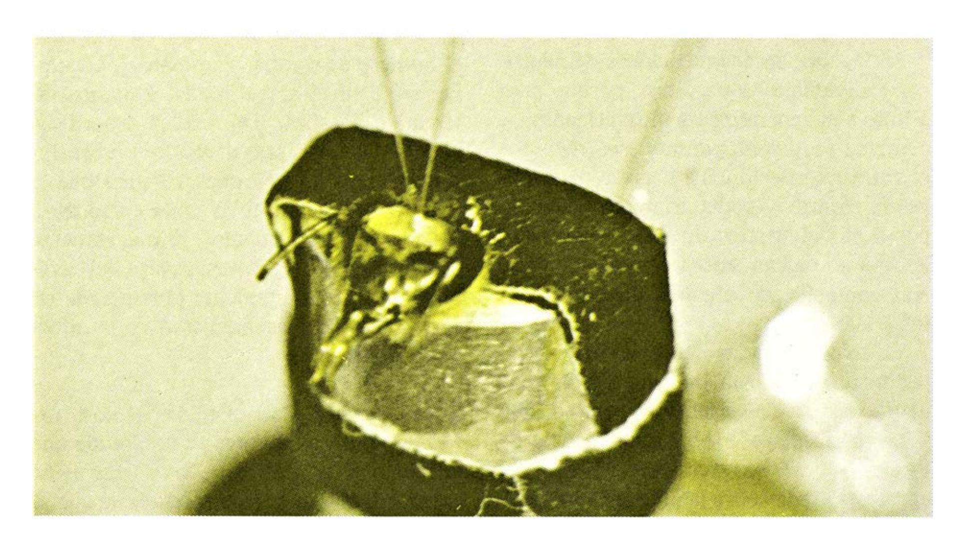
La primera etapa en un aprendizaje asociativo la constituye obviamente un mecanismo de almacenamiento fugaz de información sensorial; se trata de la fase que hemos dado en llamar almacenamiento o memoria sensorial. Si no se refuerza esa información rápidamente, desaparece. Reforzamiento de la señal visual que depende de un segundo sistema sensorial: los receptores del azúcar que se encuentran en la lengua del insecto. Si se estimulan los receptores con comida, la fase siguiente del programa de aprendizaje (almacenamiento de la información en un sistema de memoria) se realiza durante el vuelo del animal.

Parece lógico pensar que estos rápidos procesos iniciales de canalización y selección de la información se fundan en la actividad eléctrica de determinadas células nerviosas. Las grabaciones electrofisiológicas podrían ayudarnos a descubrir de qué celulas se trata. Ahora bien, ¿en qué región del cerebro deberá insertarse el electrodo de grabación? Aunque el cerebro de la abeja es relativamente pequeño (mide aproximadamente un milímetro cúbico), contiene unas 850.000 células nerviosas, demasiadas para hacer una estrategia de investigación aleatoria. Por cuya razón, hemos optado por localizar las áreas del cerebro que intervenían en los procesos de almacenamiento iniciales.

Hemos abordado el tema estudiando las respuestas reflejadas de la probóscide frente a los estímulos aportados por soluciones azucaradas. La abeja despliega su probóscide en cuanto una antena entra en contacto con una solución azucarada. Introdujimos una abeja en un pequeño tubo metálico y provocamos una vaharada de olor instantes antes de que la antena rozara la solución azucarada y que se le permitiera chupar un poco la solución con la trompa durante breves segundos. Si luego se le acercaba el olor sin el acompañamiento de la solución azucarada, la abeja extendía siempre e invariablemente la trompa. Basta uno de estos ensayos de condicionamiento para conseguir un alto nivel de respuestas.

Vimos que se podía descubrir el cerebro de la abeja mientras durara este proceso, sin que quedara afectado de modo contraproducente el comportamiento. Pudimos luego bloquear la actividad neuronal en áreas pequeñas del cerebro de modo reversible con una delgada aguja fría. El área bloqueada podía llegar a ser extraordinariamente pequeña, esto es, hasta una superficie de 250 micrometros de anchura por 150 de profundidad.

Mediante estos procedimientos llega-



HABITUACION AL OLOR por parte de las abejas. Se trata de otro método para la investigación de la formación de la memoria. Se fija la abeja dentro de un tubo, y se pone en contacto la antena del insecto con una gota de solución de sucrosa. La abeja extiende la trompa por un reflejo no condicionado. Si acercamos una vaharada de olor por breve tiempo antes de que la antena entre en contacto con la solución azucarada y se permite que la pruebe con la probóscide desplegada, el olor se convierte en un estímulo condicionado. Después de una prueba de acondicionamiento, es decir, tras una única succión de sucrosa, la mayoría de las abejas despliegan la trompa o probóscide aun cuando aparezca la vaharada sin la recompensa de comida.

mos al convencimiento de que interesaba mucho un punto de los *corpora pedunculata*, también conocidos por "cuerpos hongos". Hay dos cuerpos de éstos, cada uno de ellos repleto de fibras nerviosas. El cuerpo pedunculado posee cuatro subestructuras principales: el cáliz terminal, el pecíolo, el lóbulo frontal alfa y el lóbulo beta del fondo. Su poder principal proviene de otra estructura, el lóbulo antenal. Los histólogos del siglo XIX descubrieron que los cuerpos pedunculados eran especialmente prominentes en los insectos sociales y dedujeron que las funciones más elevadas del cerebro (aprendizaje, memoria y comportamiento social complejo) se hallaban probablemente localizadas allí.

Al dañar partes de los lóbulos antenales y de los cuerpos pedunculados por refrigeración, encontramos una correlación con lo que habríamos observado al bloquear la actividad neuronal de las abejas en vuelo libre. El efecto de las lesiones ocasionadas disminuye si aumenta el tiempo transcurrido entre el ensayo y el tratamiento. La susceptibilidad a la lesión disminuye más rápidamente en los lóbulos antenales de lo que sucede en el cáliz.

Estos resultados indican que los lóbulos antenales y parte de los cuerpos pedunculados intervienen en el proceso de transferencia de la información hasta la memoria, aunque lo hagan en distinta extensión y en diferentes periodos de tiempo. Los lóbulos antenales intervienen sólo durante un corto periodo inicial de unos tres minutos. El cáliz actúa por un periodo de mayor longitud. Puesto que puede demostrarse que puede interesarse la memoria, incluso hasta siete minutos después del enfriamiento de los cálices, tales son, probablemente, los responsables de la transferencia de información desde la memoria de corta duración hasta la memoria de larga duración. También se hallan involucrados los lóbulos alfa, conectados a los cálices por millares de fibras nerviosas.

La situación de aprendizaje con un solo intento en nuestros experimentos, nos permitió seguir el curso de la transmisión desde la memoria de corta duración hasta la memoria de larga duración. Los cursos de tiempo que encontramos en diversas estructuras del cerebro hacen que resulte muy poco probable que los procesos electroquímicos y bioquímicos implicados en la transcripción sigan una dinámica simple dentro de un área sinóptica determinada. Puesto que los procesos que bloquean la actividad neuronal en las abejas en vuelo libre, tanto para las entrenadas para un color como las que han sido entrenadas para olores, son similares, resulta que los procesos nerviosos centrales, responsables del almacenamiento inicial de la memoria, son asimismo parecidos, tanto para el color como para el olor.

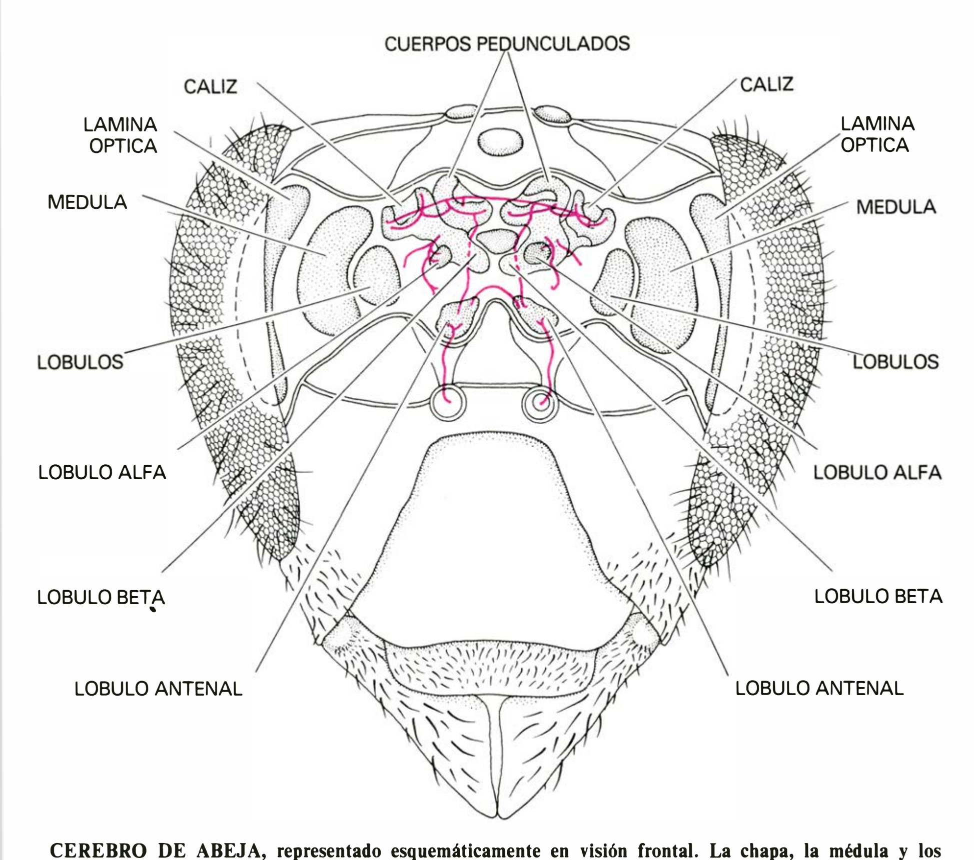
Hemos identificado áreas especiales del cerebro que intervienen en la formación de la memoria de olores. Sigue en pie la incógnita de si esas áreas están o no relacionadas con el aprendizaje visual. En cualquier caso, la duración temporal de los aprendizajes, tanto para la información visual como para la información olfatoria, resulta estrechamente afín. las grabaciones electrofisiológicas operadas en abejas inmovilizadas, y entrenadas para el olor, podrían darnos claves también para la comprensión de los mecanismos generales del aprendizaje asociativo que funcionan en las abejas en vuelo libre.

No es fácil sacar grabaciones de neuronas individuales del cerebro, porque tales células son generalmente muy delgadas. A pesar de todo conseguimos esos registros mientras el insecto tenía libertad para mover la trompa y podía estar condicionado por el olor. Nuestros estudios electrofisiológicos no acaban sino de empezar y ya se han contabilizado algunos descubrimientos generales.

La mayoría de las neuronas del cuerpo pedunculado reciben información de más de un sistema sensorial. Por ejemplo, una neurona puede responder a un destello de flash inhibiéndose (excitándose menos), al agua azucarada sobre la antena estimulándose (excitándose más) y a un olor mezclado con excitación propiamente dicha. Hemos observado combinaciones diferentes de respuestas para distintos estímulos.

Sólo unas cuantas neuronas de éstas cambian su capacidad de respuesta mientras dura la etapa de acondicionamiento. Y las que así actúan respondían ya al olor y al azúcar antes de que empezara el entrenamiento. El cambio suele consistir en un aumento del nivel de descarga espontánea de la neurona acompañado de un aumento de la sensibilidad a los estímulos de olor y agua azucarada. En algunas neuronas descubrimos una estrecha correlación entre el número de condicionamientos y el cambio en el nivel de respuesta. Estamos intentando localizar las células específicas sujetas a aprendizaje mediante tinción intracelular de las neuronas. Se diría que la mayoría se hallan en la zona en torno a los cuerpos pedunculados, lo cual refuerza la hipótesis de que es esta parte del cerebro la que está altamente involucrada en el proceso y almacenamiento de la información olfatoria.

De los trabajos acometidos por nosotros y por otros autores se desprende que la memoria asociativa de las abejas sigue las principales etapas de aprendizaje que se han establecido para los vertebrados. Es más, las abejas poseen memorias de corto y largo plazo, enteramente similares a las que se conocen en los vertebrados. Tampoco los sistemas neuronales responsables del almacenamiento del aprendizaje y la memoria en las abejas difieren en lo fundamental de los observados en los vertebrados. Por tanto, gracias al trabajo que se está llevando a cabo con las abejas, se pueden conseguir muchas probabilidades de desentrañar algunas de las reglas básicas del aprendizaje y la memoria asociativos, asimismo relevantes para los complejos sistemas nerviosos de los vertebrados.



CEREBRO DE ABEJA, representado esquemáticamente en visión frontal. La chapa, la médula y los lóbulos son los encargados principales de procesar la información visual. El cerebro medio posee los dos cuerpos pedunculados (cuerpos hongos) que están formados por interneuronas, densa y ordenadamente dispuestas. Los cuerpos pedunculados constan de los cálices, los lóbulos alfa y los lóbulos beta. Otras dos estructuras, los lóbulos antenales, constituyen los centros sensoriales primarios que reciben la información de la antena. El tramo principal del nervio que transmite la corriente de información entre los cuerpos pedunculados y los lóbulos antenales aparecen en color. Por mm³ de cerebro, posee unas 850.000 neuronas.

La tecnología de las membranas sintéticas

Las delgadas películas de polímeros entrelazados pueden separar unas moléculas de otras según su tamaño, carga u otras propiedades. Las membranas sintéticas pueden aplicarse a ciertos procesos industriales: tratamiento de residuos y desalinización

Harry P. Gregor y Charles D. Gregor

Un número, no pequeño, de procesos industriales se caracteriza simplemente por separar una cosa de otra. En minería, por ejemplo, debe extraerse el metal puro de la ganga con la que se encuentra. Cuando se desaliniza el agua de mar se obtienen dos productos: sal y agua potable. En la depuración de las aguas de alcantarillado y de muchos residuos industriales, hay que concentrar los productos de desecho separándolos de la corriente líquida para poderlos eliminar con eficacia.

Un camino muy directo para abordar estos objetivos sería la construcción de una barrera que permitiera el paso de las partículas o moléculas de un tipo y excluyera a todas las demás. Puede crearse una barrera selectiva de esta índole con una membrana sintética hecha de polímeros. Se conocen membranas de éstas desde hace muchos años, pero sólo recientemente se han superado los obstáculos que impedían su explotación comercial.

Las membranas sintéticas pueden dividir las sustancias en razón de distintas propiedades. En el proceso llamado ultrafiltración, las moléculas se separan según su tamaño, en electrodiálisis según su carga eléctrica y en la extracción con disolventes según sus diferentes afinidades por dos disolventes distintos. Las membranas que son permeables casi exclusivamente al agua se emplean en el proceso denominado ósmosis inversa. Con una distribución de varias membranas puede descomponerse una sal disuelta en agua en un ácido y una base, y uniendo un enzima a una membrana pueden provocarse reacciones químicas en las que se obtiene una separacion eficaz entre las sustancias reaccionantes y los productos de la reacción.

Quizás el caso más simple de la tecnología de membranas lo constituya la ultrafiltración; en ella las moléculas disueltas y las partículas suspendidas se seleccionan principalmente según sean o no lo bastante pequeñas para atravesar los poros de la membrana. En la filtración ordinaria se hace pasar bajo presión un líquido a través de un medio poroso, una hoja de papel o un lecho de arena, por ejemplo. La ultrafiltración no es fundamentalmente distinta, con la salvedad de que los poros son aproximadamente 1000 veces más pequeños que los de un filtro ordinario, de modo que se pueden extraer y concentrar no sólo las partículas de materia sino también las moléculas seleccionadas. Además, algunas membranas de ultrafiltración excluyen a las moléculas cargadas eléctricamente, prescindiendo de su tamaño.

Casi todas las partículas suspendidas tienen un tamaño de orden de 10^{-6} metros, es decir, un micrometro. Las partículas de polvo raramente tienen menos de un micrometro de diámetro; las bacterias poseen generalmente de cinco a 10 micrometros en su mayor dimensión. Los poros de una membrana de ultrafiltración son del orden de 10^{-9} metros, es decir, un nanómetro; el promedio de los diámetros de los poros oscila desde menos de un nanómetro a unos 10 nanómetros. De aquí que por ultrafiltración se rechacen virtualmente todas las partículas suspendidas, comprendidas las bacterias e incluso los virus, que tienen un tamaño mínimo de unos 20 nanómetros.

La ultrafiltración es esencialmente una filtración molecular; diferencia moléculas de distintos tamaños y, en términos generales, moléculas de distinto peso molecular. El agua, con un peso molecular de 18 y un diámetro efectivo de unos 0,2 nanómetros, puede pasar libremente a través de todas las membranas de ultrafiltración. Los azúcares sencillos, tales como la glucosa (que tiene un peso molecular de 180 y un tamaño de medio nanómetro, aproximadamente) pueden atravesar también la mayoría de las membranas de ultrafiltración. Por otra parte, las proteínas y otras macromoléculas biológicas suelen quedar excluidas, ya que frecuentemente tienen un peso molecular próximo a 100.000 y un tamaño de varios nanómetros. Si se escoge una membrana con la medida de los poros cuidadosamente controlada, se pueden incluso separar moléculas con pocas diferencias de tamaño, tales como la sacarosa (peso molecular 342) y la cianocobalamina, o la vitamina B-12 (peso molecular 1300).

Una membrana de ultrafiltración se hace o "moldea" a partir de una solución de polímeros de cadena larga, como el acetato de celulosa disuelto en acetona. La solución es viscosa (bastante parecida a la miel): se esparce con la hoja de un cuchillo de manera que forme una capa delgada. Luego se deja evaporar parte del disolvente, y el polímero empieza a sobresalir de la solución, formando una matriz delgada y semisólida. Finalmente, la membrana se sumerge en otro disolvente, puede ser el agua, que precipita rápidamente todo el polímero restante. Este enfriamiento brusco por inmersión es el responsable de la formación de los poros de la membrana, mientras que la rápida precipitación conduce a la coagulación del polímero. El tamaño de los poros viene determinado principalmente por la rapidez del secado parcial y por la elección de los disolventes. Aunque los poros no son nunca completamente uniformes, su diámetro promedio puede controlarse bastante bien.

Las membranas fabricadas de esta manera son bastante delgadas, generalmente de cinco a 30 micrometros. Ya que deben soportar una cierta presión, se les puede proporcionar un soporte mecánico, me-

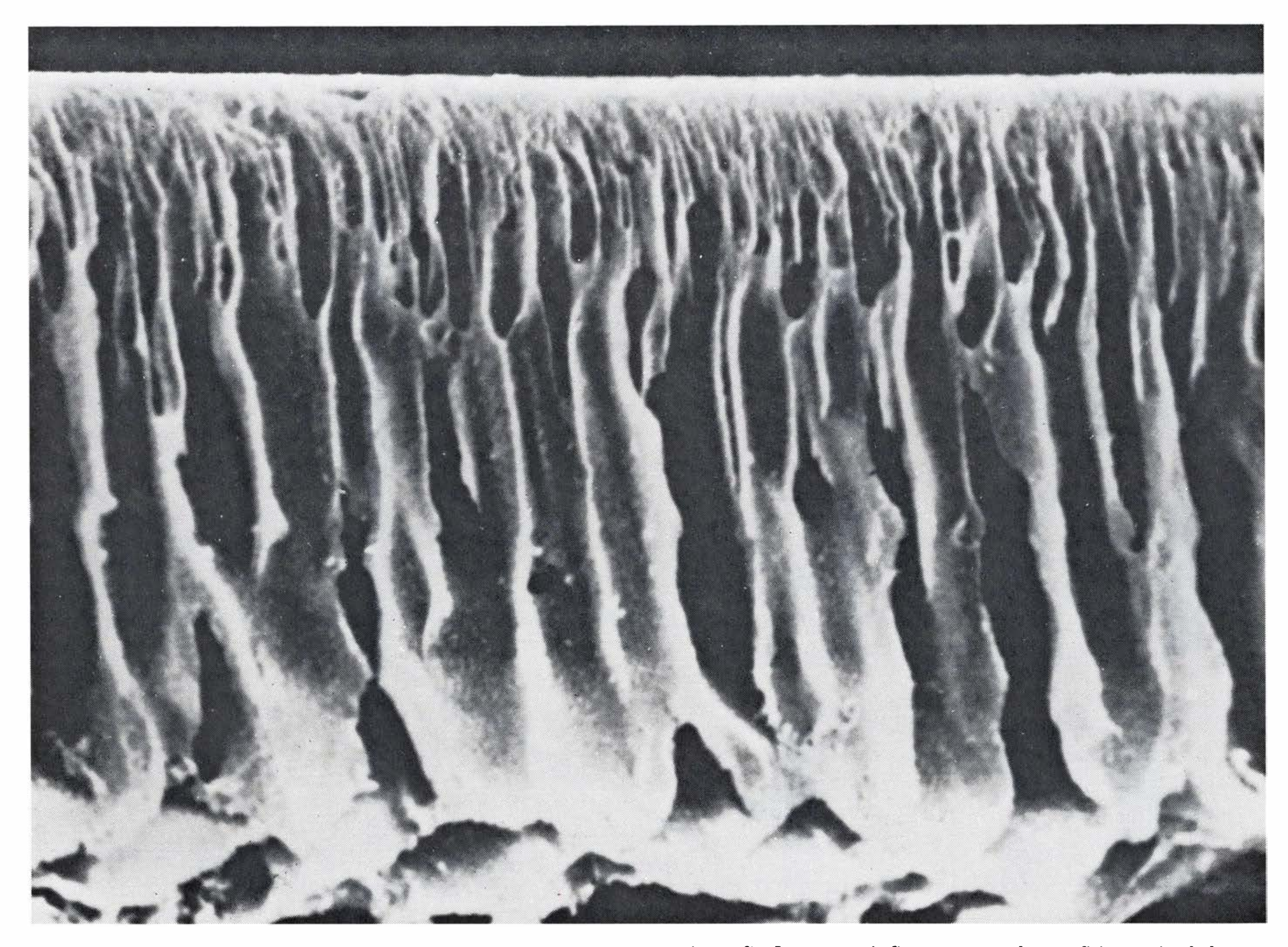
diante un sustrato con poros microscópicos. Puede moldearse una membrana sobre una plancha de vidrio, y arrancarla y montarla después en un sustrato apropiado. Sin embargo, hoy en día es más frecuente moldear la membrana directamente en un soporte poroso, como plástico, papel o ropa.

Probablemente el primer aparato sintético de ultrafiltración fue el realizado por el químico y biólogo alemán Moritz Traube en 1870. Traube precipitó un gel de ferrocianuro cúprico en los poros macroscópicos de un florero no vidriado. Por los años 30 ya se habían desarrollado membranas de polímeros de ultrafiltración con poros de varios tamaños. Hay que destacar el trabajo realizado por su contemporáneo William J. Elford, que fabricó membranas de acetato de celulosa y nitrato de celulosa; a él se le debe también la sistematización de la mayoría de los principios de la ultrafiltración.

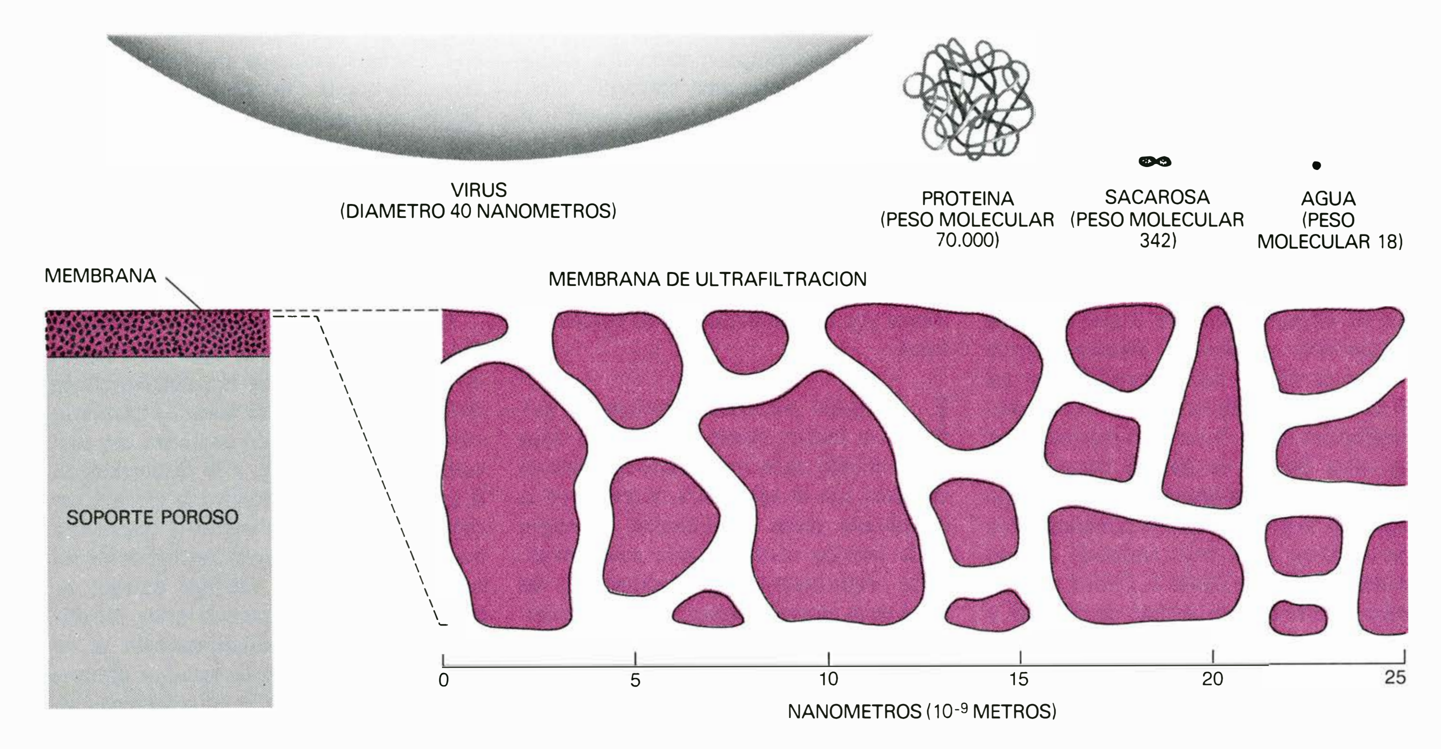
La teoría de la ultrafiltración puede explicarse por un simple modelo en el que se interpone un membrana entre dos líquidos, siendo uno de ellos agua pura y el otro agua con alguna sustancia disuelta. Se puede suponer inicialmente que la membrana es totalmente permeable al agua y completamente impermeable al soluto. La magnitud y dirección del flujo a través de la membrana vienen regidas entonces por una combinación de cuatro factores.

La presión osmótica constituye el primer factor; tiende a dirigir el agua desde el lado de la membrana que está en contacto con el agua pura, a través de la membrana, hasta la disolución. El origen de la presión osmótica sólo puede explicarse rigurosamente introduciendo las leyes de la termodinámica y, en particular, el concepto de entropía; y puede entenderse intuitivamente como una tendencia del disolvente (en este caso el agua) a igualar su concentración en todo el volumen, como si la membrana no estuviera presente. Ya que la concentración del agua es mayor en el agua pura que en una disolución, sus moléculas tienden a pasar, a través de la membrana, a la disolución. El valor de la magnitud de la presión osmótica equivale a la diferencia entre las concentraciones del agua a ambos lados de la membrana.

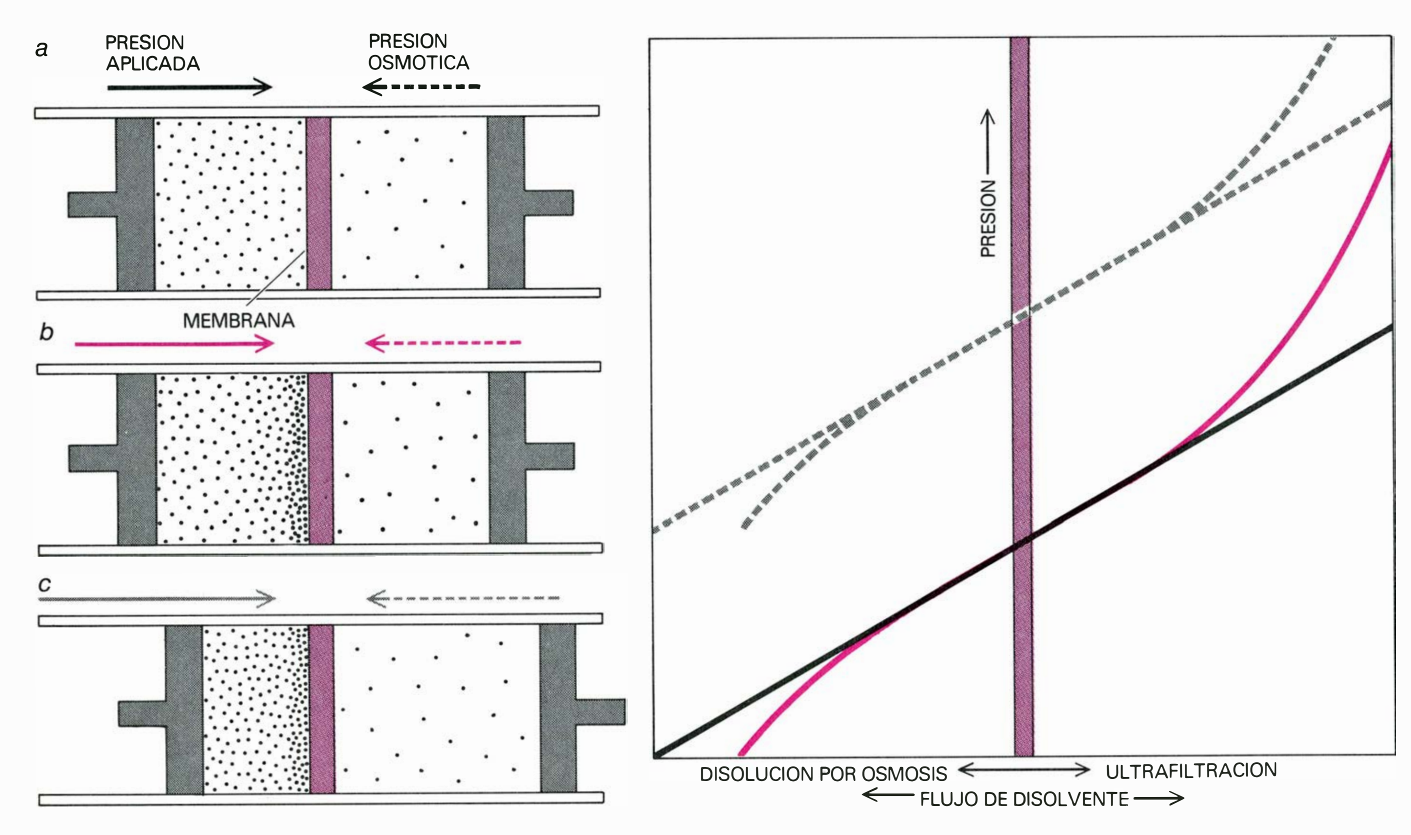
Si este sistema de membrana se dejara funcionar por sí solo, el agua pura fluiría de manera que diluiría la solución, efecto opuesto al deseado. Por tanto, es necesario contrarrestar la presión osmótica con una contrapresión aplicada a la disolución. Si la presión aplicada excede a la presión osmótica, el agua abandonará la disolución quedando así una mayor concentración de soluto. La velocidad del flujo de agua es proporcional a la diferencia entre las dos presiones. Esta velocidad también se ve afectada por otros dos factores: es directamente proporcional a la permeabilidad de la membrana al agua e inversamente pro-



MEMBRANA POROSA vista en una sección transversal; en esta microelectrografía aparece ampliada unas 500 veces. La membrana se formó a partir de una solución de polímeros que a continuación se extendió formando una capa delgada y se dejó secar parcialmente, sumergiéndola después en agua. El agua extrajo rápidamente el disolvente que quedaba, con lo cual precipitó el polímero, formando los poros que a modo de canales verticales aparecen en la micrografía. Los poros más finos cercanos a la superficie superior de la membrana se formaron dejando que las capas superficiales se secaran exhaustivamente, de suerte que apareciera una "piel" sobre la solución del polímero. Esta membrana, fabricada por la Amicon Corporation, se diseñó para ser utilizada en ultrafiltración, proceso por el que se puede eliminar de un líquido no sólo las partículas de materia suspendidas sino también muchas moléculas disueltas.



UNA MEMBRANA DE ULTRAFILTRACION tiene poros de dimensiones moleculares. La membrana en sí es una película de 5 a 30 micrometros de grosor moldeada sobre un sustrato poroso mucho más grueso, como papel o ropa, que proporciona el soporte mecánico. Las membranas pueden fabricarse con poros que tengan un diámetro promedio que oscile de 1 a 10 nanómetros (siendo un nanómetro la mil millonésima parte del metro). La membrana que se muestra aquí *(abajo a la derecha)* en sección transversal esquemática tiene poros de un nanómetro de diámetro aproximadamente. El agua pasa fácilmente a través de estos poros, así como las moléculas pequeñas, las de sacarosa, por ejemplo. Sin embargo, la mayor parte de las proteínas quedan excluidas, como también quedan excluidos los virus, que son mucho mayores, y otros microorganismos. Una membrana de ultrafiltración llega a separar macromoléculas.

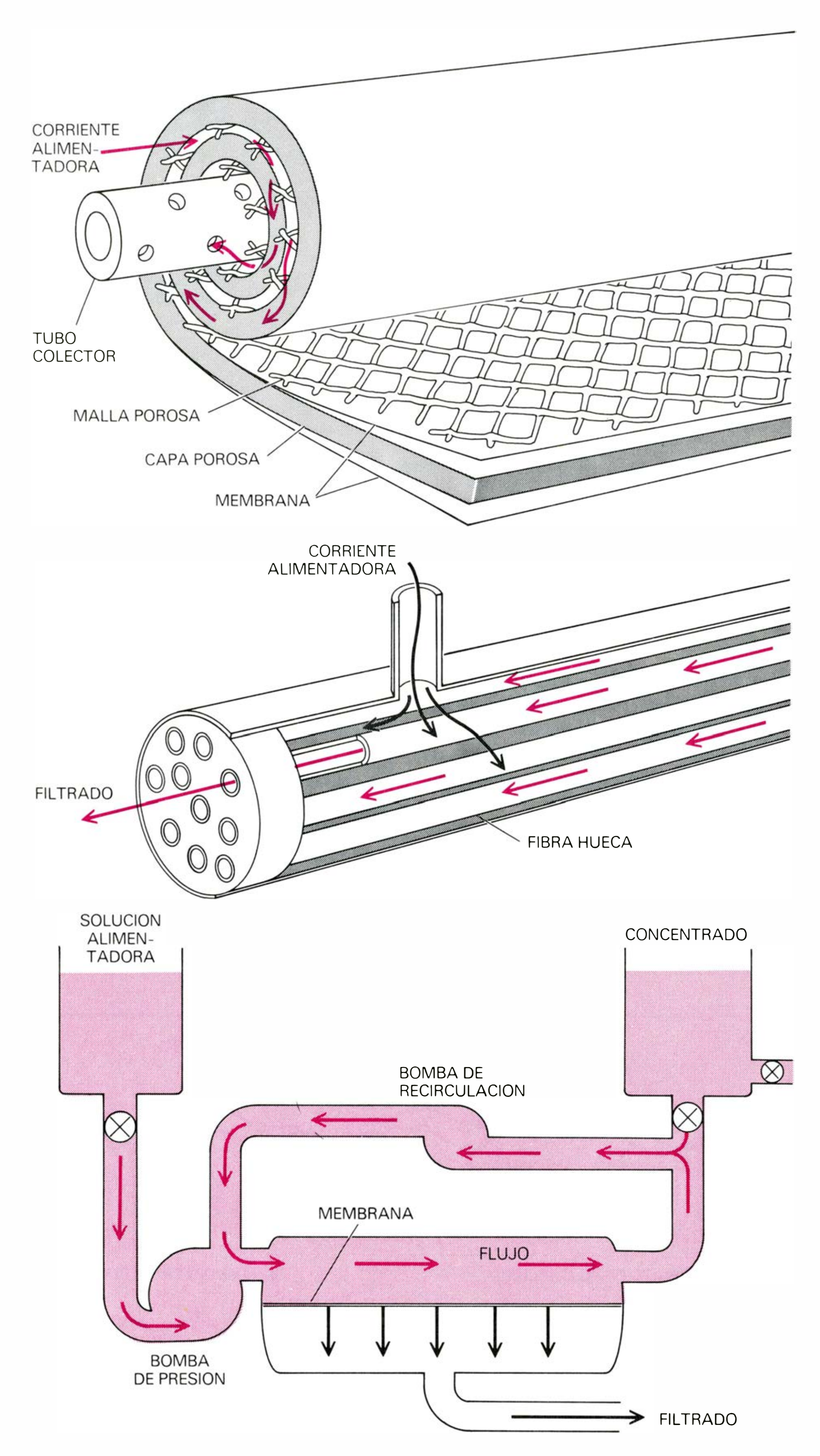


EL FLUJO DEL AGUA o de otro disolvente a través de una membrana de ultrafiltración depende, en parte,de las propiedades intrínsecas de la membrana y, en parte, de la presión. Una diferencia de concentración produce una presión osmótica que tienda a dirigir el agua pura a través de la membrana hacia la solución más concentrada, desencadenando un flujo en la dirección opuesta a la deseada. Si ha de tener lugar la filtración, esta presión debe ser contrarrestrada por una presión aplicada más grande. En el caso ideal (*a, curva en negro*), el flujo a través de la membrana es simplemente proporcional a la diferencia entre la presión aplicada y la presión osmótica. En la práctica, cualquier flujo apreciable provoca una acumulación de moléculas de soluto a lo largo de la superficie de la membrana orientada hacia la zona de mayor concentración (*b, curva coloreada*); esta condición, denominada polarización de la concentración, incrementa la presión osmótica y, por tanto, aumenta la presión aplicada que se necesita para mantener un flujo dado. Además, si no se diluye continuamente la solución alimentadora, su concentración va aumentado con la filtración (*c, curva en gris*), incrementando todavía más la presión osmótica existente.

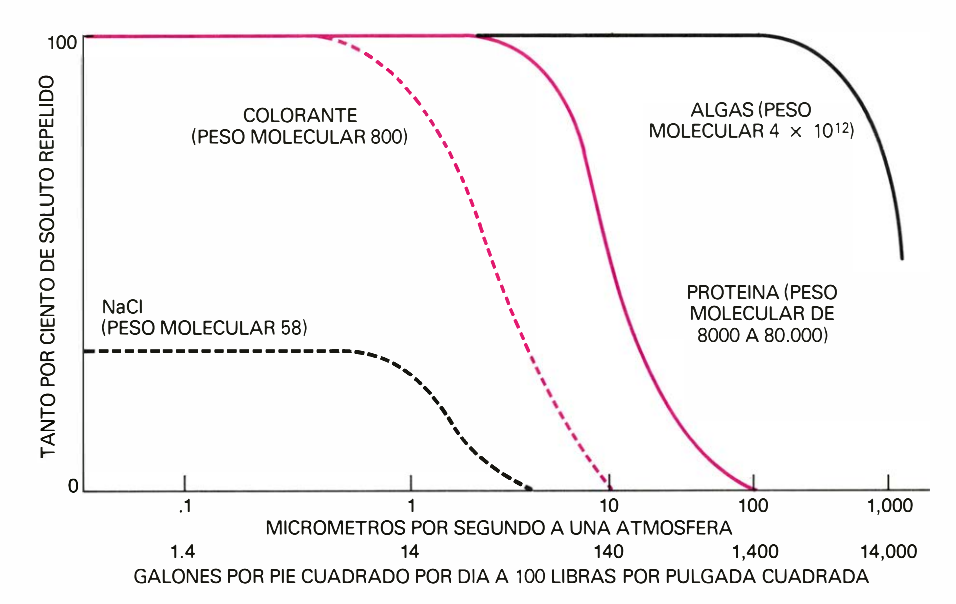
porcional al grosor de dicha membrana. Debido a que incluso una solución moderadamente concentrada puede originar una presión osmótica de algunos cientos de libras por pulgada cuadrada (una libra por pulgada cuadrada equivale a 6,89 kilonewton por metro cuadrado), a menudo se necesitan aplicar altas presiones en ultrafiltración.

Una molécula que resulte demasiado grande para atravesar los poros de una membrana será totalmente repelida por ultrafiltración; pero incluso las moléculas más pequeñas, aquellas para las cuales la membrana es permeable, pueden concentrarse hasta cierto punto. El flujo de estos solutos viene determinado predominantemente por el proceso de difusión, que describe el movimiento aleatorio de las moléculas y que es independiente de la presión. Las moléculas de un soluto para el cual la membrana sea permeable se difunden a través de ella en ambas direcciones, pero el movimiento neto es hacia la región de concentración más baja. El valor de la magnitud del flujo de soluto es directamente proporcional a la diferencia de concentraciones e inversamente proporcional al grosor de la membrana. Ya que el movimiento del disolvente es también inversamente proporcional al grosor de la membrana, la repulsión efectiva de un soluto para el cual la membrana sea permeable no depende del grosor de la membrana. Varía solamente con la presión aplicada, ya que las altas presiones aumentan el flujo de agua sin afectar al flujo de soluto.

En este análisis se ha supuesto que tanto la concentración del agua como la del soluto se mantiene uniforme en todo el volumen a cada lado de la membrana. En realidad esta condición no puede darse. Junto a la membrana hay siempre una capa límite de disolución no agitada, estancada, de donde se va extrayendo constantemente el agua y donde, por tanto, los solutos rechazados se van acumulando en concentraciones excepcionalmente altas. Este efecto, que se denomina polarización de la concentración, aumenta la presión osmótica que se opone a la filtración, ya que la diferencia "percibida" de concentración del disolvente es la de los fluidos más cercanos a la membrana. El grado de polarización de la concentración aumenta proporcionalmente al flujo de disolvente, y también es proporcional a la presión aplicada. Como resultado, existe en realidad una velocidad máxima de flujo por unidad de área para cada membrana; más allá de la cual, cualquier incremento ulterior de la presión aplicada no produce un flujo de disolvente más grande sino que sólo provoca un aumento



APARATO DE ULTRAFILTRACION que puede suministrar un flujo de solución concentrada de un lado a otro de la superficie de la membrana para debilitar la obstrucción y reducir la polarización de la concentración. En una de las configuraciones (*arriba*) dos membranas separadas por una capa porosa están enrolladas junto con un espaciador de malla de plástico alrededor de un tubo poroso. Se bombea la solución a través del espaciador y el líquido filtrado, después de atravesar la membrana, circula en espiral a través de la capa separadora hacia el tubo central del "brazo de gitano", donde se recolecta. Las membranas sintéticas también pueden arrollarse formando delgadas fibras huecas (*en medio*) con un diámetro interno de unos 40 micrometros. Las fibras se disponen formando una especie de manojo y la solución alimentadora se hace circular redeándolas de modo que el líquido filtrado salga por los extremos de las fibras. En cualquier célula de ultrafiltración se necesita una bomba para proporcionar la presión que dirige la filtración; a menudo se emplea una segunda bomba para recircular el concentrado y conseguir un flujo en la dirección al filtrado.



PERMEABILIDAD DE LAS MEMBRANAS a las moléculas de varios tamaños; está relacionada con el tamaño del poro, y éste a su vez se mide por la velocidad a la cual el agua fluye a través de la membrana bajo una presión prefijada. Las membranas más compactas permiten que pasen volúmenes considerables de cloruro sódico, y las menos compactas bloquean hasta células de algas. Entre estos extremos, algunas membranas repelen sólo las macromoléculas; otras excluyen las moléculas de un tamaño considerable.

de la presión osmótica. La polarización de la concentración puede causar también otros fenómenos no deseables, como la precipitación de sólidos en la superficie de la membrana; por cuya razón se evita limitando la presión aplicada.

El flujo a través de una membrana de ultrafiltración puede medirse en función del volumen transmitido por unidad de área de la membrana en un período dado y bajo una presión aplicada, asimismo determinada, o en función de la velocidad impartida de la corriente por una diferencia de presión dada. Como era de esperar, el flujo crece al aumentar el tamaño de los poros. La unidad práctica es el galón por pie cuadrado y por día (un galón norteamericano son 3,78 litros; el pie cuadrado vale 0,0929 metros cuadrados) bajo una presión aplicada de 100 libras por pulgada cuadrada; la unidad métrica es el micrometro por segundo bajo una diferencia de presión de una atmósfera.

Resultaría difícil medir directamente el tamaño de los poros de una membrana, pero este tamaño puede estimarse calculando la velocidad del flujo bajo condiciones prefijadas, que viene determinada en gran parte por el tamaño del poro. El flujo medido puede calibrarse observando la repulsión de moléculas de diversos tamaños en una membrana con una velocidad de flujo conocida. Las moléculas usadas normalmente para la prueba son la albúmina de suero de buey (una proteína de peso molecular 75.000), el colorante eritrosina (peso molecular 800) y la sacarosa (peso molecular 342). Una membrana "compacta", con un flujo de un micrometro por segundo a una atmósfera (o equivalente a 14 galones por pie cuadrado y por día a 100 libras por pulgada cuadrada) rechaza casi completamente las moléculas más grandes y excluye algunas del azúcar. Una membrana de flujo medio, aproximadamente tres micrometros por segundo a una atmósfera, rechaza el colorante y la proteína, pero permite pasar la mayoría de las moléculas del azúcar. Una membrana caracterizada como "rápida" (con un flujo superior a los 50 micrometros por segundo) permite pasar incluso a la proteína, pero excluye las partículas de materia tales como los microorganismos.

HEPARINA

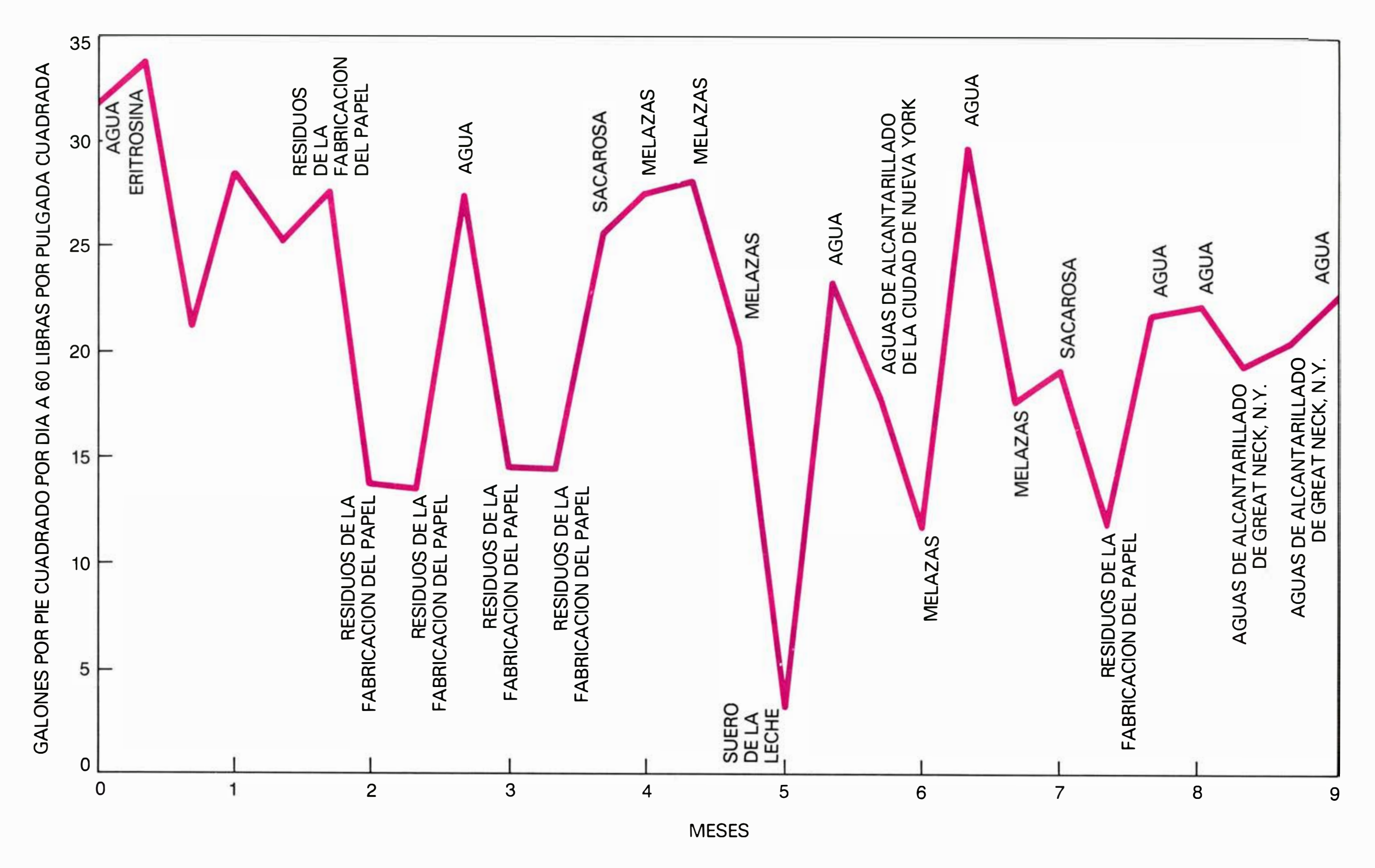
SULFONATO DE POLIESTIRENO

SULFONATO DE POLIVINILO

LOS POLIMEROS QUE EVITAN LA OBSTRUCCION incorporan a su estructura el grupo sulfonato (SO_3^-), que les confiere una intensa afinidad por el agua e impiden que se adhiera gran parte de la materia extraña. La mayor parte de las partículas obstructoras son hidrofóbicas, es decir, repelen el agua; la mayoría de ellas tienen también carga eléctrica negativa. La heparina es un agente antiobstructor que mantiene húmedos los vasos sanguíneos. Los grupos químicos activos de la heparina son los sulfonato.

En la filtración ordinaria se bombea simplemente la corriente alimentadora con sus materiales suspendidos y disueltos, a través del medio filtrante, que a menudo se trata de un lecho de arena. Las partículas de la corriente finamente divididas se adhieren al filtro y unas a otras, formando una capa de lodo que se conoce por su nombre alemán "schmutzdecke". Debido a que el schmutzdecke se vuelve pronto impermeable, el filtro debe reemplazarse o lavarse periódicamente para mantener una velocidad de flujo satisfactoria. En la ultrafiltración se emplea una técnica distinta. La solución alimentadora se hace circular continuamente de un lado a otro de la superficie de la membrana, procedimiento que ayuda a evitar la obturación, además de reducir la polarización de la concentración.

El aparato de ultrafiltración se puede montar de varias maneras. Por ejemplo, pueden montarse hojas paralelas de membrana en un marco, o se puede recubrir de membrana un tubo poroso. Una configuración eficaz es el llamado "brazo de gitano" que se construye enrollando membranas y espaciadores alrededor de un tubo poroso que sirve de colector del líquido filtrado. Las membranas se pegan a cada lado de una hoja de papel o ropa porosa, entonces la doble membrana se enrolla con un espaciador de malla de plástico alrededor del tubo colector. La so-



RESISTENCIA A LA OBTURACION de una membrana fabricada a partir de un polímero con grupo sulfonato, valorada exponiendo continuamente una única membrana a una serie de líquidos prueba durante meses. Varios de éstos, como los líquidos residuales de la fabricación de la pasta de papel, las melazas, el suero de la leche y las aguas del alcantarillado, representan una dura prueba a la eficacia filtradora y obturarían las membranas ordinarias en minutos u horas. La membrana sulfonada permanece limpia incluso al final del período de prueba. Las velocidades de flujo indicadas se midieron al final de cada prueba. Las variaciones de velocidad son debidas quizá menos a la obturación de los poros que a las diferencias de presión osmótica inducidas por cada solución.

lución alimentadora es bombeada a través del espacio comprendido entre las capas de doble membrana, y el líquido filtrado circula en espiral a través del medio poroso hacia el tubo central. Este diseño es considerablemente más barato que los filtros de hojas paralelas o de tubo recubierto y asegura una fuerte contracorriente a lo largo de la superficie de la membrana en el lado concentrado.

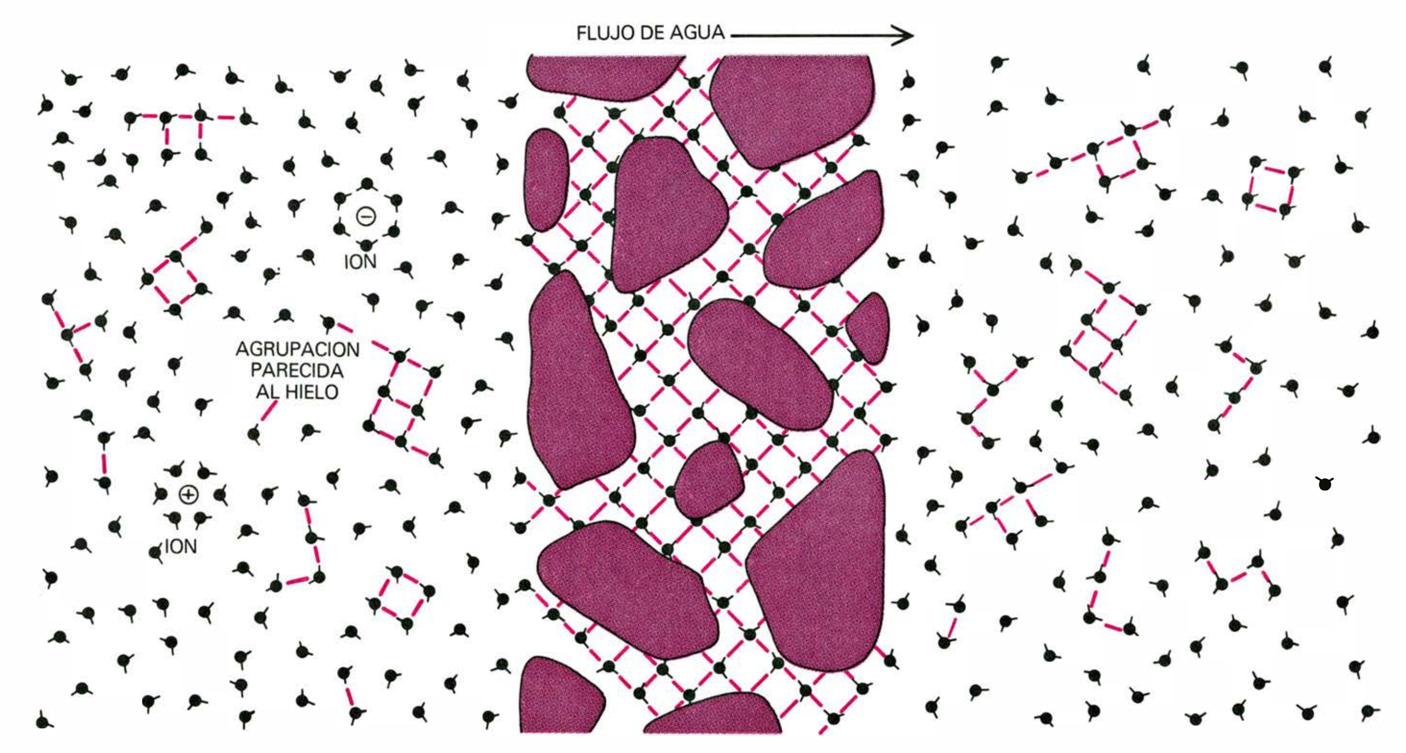
En otra configuración favorecida la membrana se enrolla formando una fibra hueca, normalmente no más gruesa que un cabello humano. Debido a que las fibras tienen gruesas paredes en comparación con sus dimensiones completas, normalmente puede soportar presiones de hasta 1500 libras por pulgada cuadrada. El coste es más o menos el mismo que para el modelo en espiral, pero la geometría del modelo de fibra hueca lo hace susceptible a la obturación. Si la corriente alimentadora pasa a través del centro hueco las fibras tienden a obturarse con partículas de materia, y si se hace circular por fuera de las fibras la contracorriente no es suficiente.

La ultrafiltración se desarrolló en los años 20 pero durante 4 décadas este proceso se empleó muy raramente fuera del laboratorio, debido sobre todo a un único problema: la obturación de las membranas. Cuando una membrana se exponía a una corriente alimentadora que contenía sólidos en suspensión, las partículas se adherían a la superficie y obstruían los poros. La velocidad del flujo descendía rápidamente y en horas o en minutos el filtro dejaba de funcionar.

Hacia 1960 uno de nosotros (Harry P. Gregor) empezó un estudio sistemático de la obturación de las membranas con el ánimo de preparar membranas que la resistieran. Pronto se hizo evidente que los problemas de obturación más serios eran los causados por materiales tales como las arcillas, partículas grasientas y proteínas, todas las cuales tienen grandes áreas superficiales que son hidrofóbicas, o en otras palabras, que repelen el agua. Cuando una sustancia hidrofóbica se encuentra en un ambiente acuoso, puede reducir su energía total recortando el área expuesta al agua; dos partículas hidrofóbicas, por ejemplo, tienden a agruparse expulsando el agua del espacio que las separa con lo cual reducen su área superficial expuesta. De la misma manera, una partícula de este tipo puede adherirse a la superficie de la membrana debido a que así elimina cierta fracción de las interacciones repulsivas con el agua que la rodea. La mayoría de los materiales obstructores también tienen carga eléctrica negativa, y el enlace de hidrógeno en el que están implicadas estas cargas también puede contribuir a la obturación. En este tipo de enlace la carga positiva débil de un átomo de hidrógeno de la superficie de la membrana atrae a un grupo cargado negativamente de la partícula obstructora.

Una manera obvia de evitar las interacciones hidrofóbicas es crear una membrana que sea extremadamente hidrofílica, es decir, que tenga una gran afinidad por el agua. Este material permanecerá húmedo incluso en presencia de las partículas hidrofóbicas, de modo que éstas no podrán adherirse a su superficie expulsando el agua.

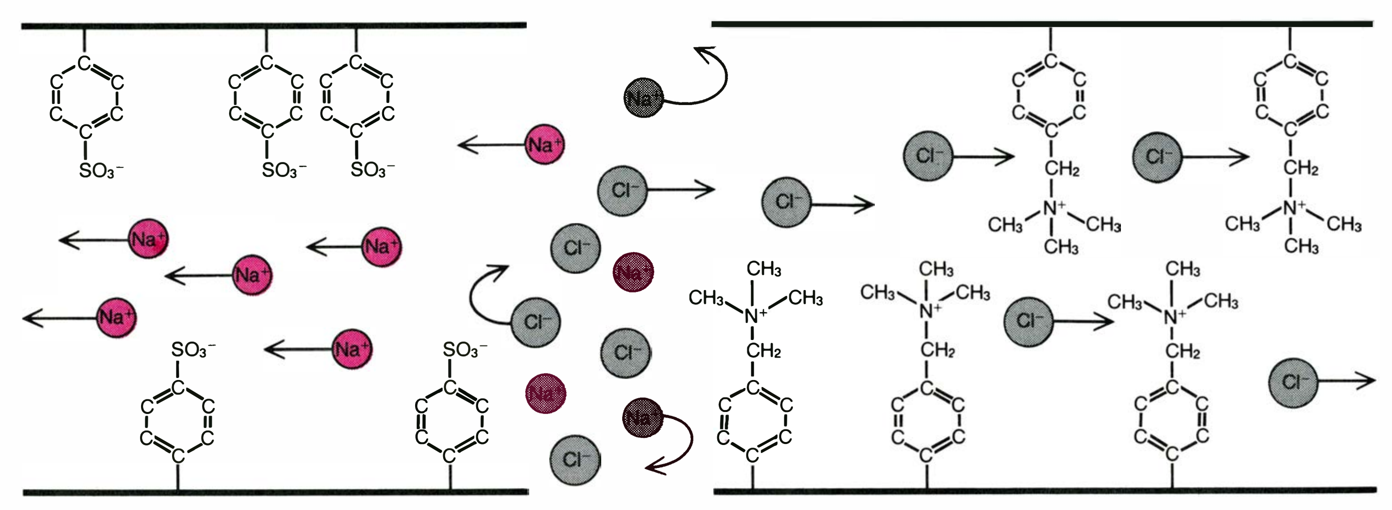
Los polímeros más hidrofílicos conocidos son los que tienen en su superficie el grupo sulfonato, SO_3^-. Al igual que el ácido sulfúrico del que derivan, los polímeros con grupos sulfonatos presentan una gran afinidad por el agua. Cada grupo



LA OSMOSIS INVERSA se parece a la ultrafiltración en que se bombea el agua bajo presión a través de la membrana, mientras los solutos van siendo excluidos, pero el mecanismo es muy distinto. La membrana de ósmosis inversa no tiene poros definidos, sólo unos espacios entre las fibras del polímero que pueden absorber un pequeño volumen de agua. Allí el agua adopta un estado parecido al del hielo (*en color*) en el que las moléculas están dispuestas de forma ordenada y se mantienen unidas por enlaces de hidrógeno. (Incluso en el líquido se van formando y desintegrando constantemente pequeñas agrupaciones en estado "de hielo".) La geometría del estado de hielo es, en realidad, tetraédrica; cada átomo de oxígeno está rodeado por otros cuatro a igual distancia, pero aquí se muestra proyectado sobre un plano. Bajo presión, las moléculas de agua se incorporan a la configuración de estado de hielo en el lado concentrado de la membrana (*izquierda*) y se van desprendiendo al otro lado (*derecha*) de modo que existe un flujo efectivo de agua a través de la membrana. Otras moléculas y partículas son repelidas, incluyendo no sólo las que son demasiado grandes para pasar a través de la membrana sino también las pequeñas moléculas que no pueden adoptar la estructura del estado de hielo. Se excluyen los iones al ir protegidos por una capa de agua que rompería el enrejado del estado de hielo.

sulfonato está normalmente rodeado por muchas moléculas de agua, que no pueden ser desplazadas por las partículas hidrofóbicas. El grupo sulfonato es también incapaz de establecer un enlace de hidrógeno; en realidad su carga negativa repele las partículas obstructoras, cargadas negativamente. La carga negativa del sulfonato se mantiene incluso en presencia de grandes concentraciones de iones hidrógeno (o sea a *p*H bajo), porque el sulfonato es la base conjugada de un ácido muy fuerte. Además, el grupo sulfonato resiste la desactivacion por iones positivos, como los de los metales pesados. Incluso las sales de plomo y de bario de los polímeros con grupos sulfonato permanecen solubles en agua, indicando así que los grupos sulfonato están todavía hidratados, mientras que la mayoría de las otras sales de plomo y de bario son casi totalmente insolubles.

La elección de los polímeros sulfonados como material base para la fabricación de membranas no obturables se vio apoyada



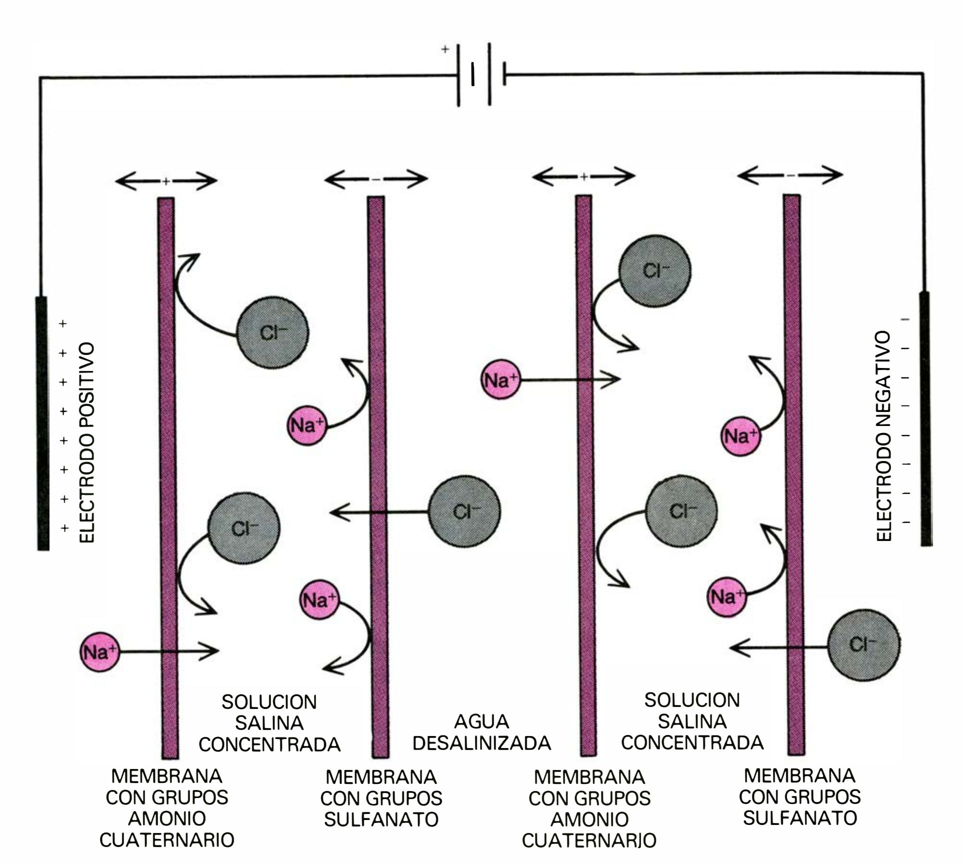
LAS MEMBRANAS CARGADAS distinguen entre iones en solución según su carga eléctrica. Aquí se muestran esquemáticamente los poros de dos membranas cargadas. La membrana cargada negativamente (*izquierda*) tiene grupos sulfonato a lo largo de las paredes de los poros y se parece a una membrana de ultrafiltración no obturable excepto en que la densidad de grupos cargados es mayor. Las membranas con grupos sulfonato permiten que pasen los iones positivos, como los de sodio (Na^+) pero bloquean los iones negativos, como los cloruros (Cl^-). La membrana cargada positivamente (*derecha*) presenta la selectividad opuesta. Las cargas positivas son aportadas por un derivado del ion amonio cuaternario (NH_4^+). Los dibujos han sido realizados por G.V. Kelvin.

por un ejemplo que se tomó de la propia naturaleza: la sustancia coloidal heparina, que es secretada por el revestimiento de los vasos sanguíneos. Se piensa que la función de la heparina es mantener las paredes de los vasos sanguíneos humedecidas y sin sedimentos. Su estructura es la de un polímero con grupos sulfonatos.

Se desarrollaron procedimientos para fabricar membranas de polímeros basados en el estireno con grupos sulfonatos expuestos en la superficie de la membrana y en los poros. En general, estos procedimientos implicaban la formación de una delgada película de polímero sulfonado, incorporando luego un agente químico que entrelazara estas moléculas de cadena larga.

Cuando se probaron estas membranas sulfonadas con soluciones alimentadoras que causaban normalmente graves obstrucciones, los resultados fueron extraordinarios. El procedimiento típico de la prueba consistía en montar una membrana en el aparato adecuado y hacerlo funcionar con una única solución hasta que su rendimiento quedara seriamente afectado. Después se instalaba una nueva membrana para la siguiente solución prueba. Con las membranas sulfonadas fue posible exponer una única membrana a sucesivos líquidos obstructores durante un período de varios meses con una pérdida insignificante de eficacia en el filtrado. Por ejemplo, se probó una membrana durante nueve meses con una sucesión de líquidos que incluían disoluciones salinas, colorantes, residuos procedentes de la fabricación de papel, melazas, suero de la leche procedente de la fabricación de quesos y aguas residuales. El flujo a través de la membrana osciló de una solución a otra, pero la variación no fue causada probablemente por la obstrucción sino por las diferencias de presión osmótica. Al final de esta segunda prueba la membrana estaba aún limpia y brillante.

La resistencia a la obturación que presentan los polímeros con ácido sulfónico pueden hacerlos útiles en otros campos que no sea la tecnología de membranas. Uno de nosotros (Harry P. Gregor) preparó pantallas y tubos recubiertos de estos polímeros y demostró que permanecían relativamente limpios incluso bajo condiciones adversas. Un revestimiento de este tipo podría inhibir la adhesión de microorganismos marinos en los aparatos sumergidos en el mar durante largos períodos de tiempo. Una aplicación posible de estos revestimientos quizá se diera en los aparatos que obtienen la energía de los gradientes de temperatura del mar. La obturación de las superficies de transferencia de calor de estos aparatos reducen considerablemente su eficacia.



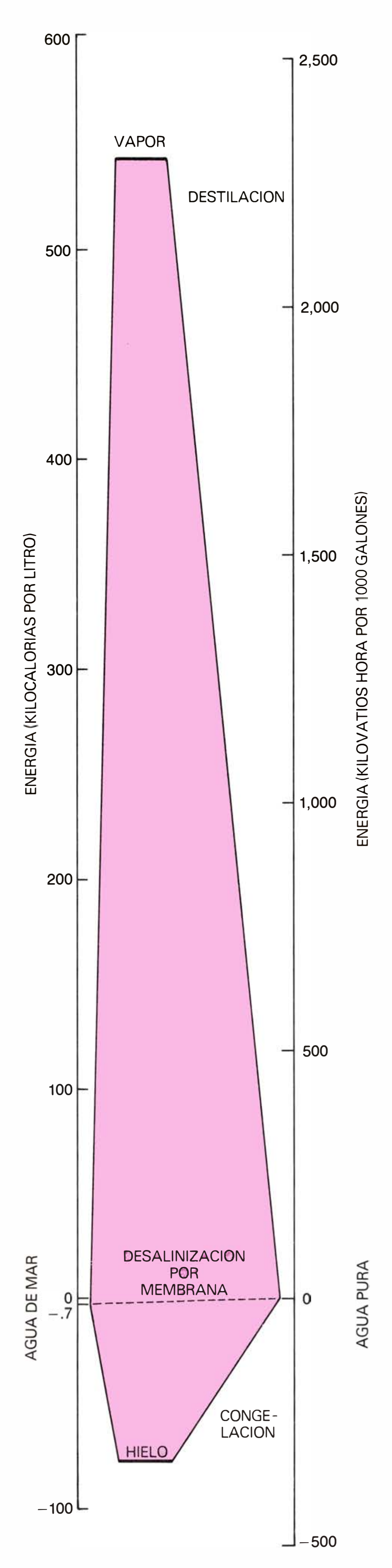
EN ELECTRODIALISIS se emplean membranas de carga fija para extraer el agua pura de una solución salina. Este proceso no está dirigido por la presión aplicada (como la ultrafiltración y la ósmosis inversa) sino por un campo eléctrico. Las membranas se identifican en la parte superior según la carga iónica que dejan pasar. Los iones cloruro son atraídos por el electrodo positivo y, por tanto, se desplazan hacia la izquierda, atravesando todas las membranas sulfonadas que encuentran a su paso, pero quedando bloqueados por las que poseen iones amonio cuaternario. Asimismo, los iones sodio se desplazan a la derecha hasta que quedan frenados por una membrana con grupos sulfonato. El agua pura se concentra en la cavidad central y la solución salina en los espacios comprendidos entre los dos pares de membranas periféricas.

Además de proporcionar resistencia a la obturación, las cargas fijas de los grupos sulfonato confieren otra propiedad a las membranas: aumentan la repulsión de las sales. Las membranas sin carga, incluso las que tienen poros bastante pequeños, pueden excluir sólo en parte las sales que existen en la solución como iones positivos y negativos. Sin embargo, en las membranas sulfonadas las cargas negativas fijas generan un campo eléctrico que repele los iones de igual carga; de este modo los excluye de los poros de la membrana. Ya que debe mantenerse la neutralidad eléctrica de la solución, los iones positivos quedan también bloqueados.

La repulsión de las sales por las membranas de ultrafiltración cargadas es eficaz siempre que la disolución salina sea relativamente diluida. En un disolución salina concentrada hay suficientes iones positivos disponibles para rodear y apantallar las cargas negativas de los grupos sulfonatos, permitiendo que pasen ambos tipos de iones. Debido a esta protección, la ultrafiltración no es una tecnología práctica para la desalinización del agua de mar que contiene unos 35 gramos de sal por litro. Sin embargo, con las aguas salobres, que contienen unos 5 gramos de sal por litro, se puede lograr una reducción de salinidad del 80 por ciento. Y lo que aún es más, los iones metálicos divalentes, tales como el ión calcio Ca^{++} y el ión magnesio Mg^{++}, se separan preferentemente por estas películas cargadas. Estos iones son los que más contribuyen a la "dureza" del agua y, por tanto, estas membranas pueden emplearse para ablandar el agua.

Si se mengua progresivamente el tamaño de los poros de una membrana, se llega a un punto en el que la membrana es permeable al agua pero a casi nada más. Se puede extraer el agua de una solución haciéndola pasar a través de una membrana de este tipo bajo presión, fenómeno denominado ósmosis inversa. Se podría pensar que la ósmosis inversa no es más que una ultrafiltración a escala extremadamente pequeña, pero de hecho su mecanismo es completamente distinto. Una señal de que existe tal diferencia es el hecho de que ni siquiera las moléculas tanto o más pequeñas que el agua atraviesen la membrana. En realidad, la membrana no tiene poros definidos, e incluso el agua no fluye a su través en el sentido usual; en lugar de ello el agua está dentro de la estructura de la membrana en un estado parecido al de hielo.

En cualquier momento, en el agua



líquida aproximadamente la mitad de las moléculas están formando agrupaciones que tienen la misma estructura ordenada que un cristal de hielo, hecho éste demostrado por primera vez en 1933 por J.D. Bernal y R.H. Fowler. En estas agrupaciones, cada molécula de agua se coloca de manera que el átomo de oxígeno ocupe el vértice de un tetraedro y que cada par de moléculas de agua esté unido por un enlace de hidrógeno. En el hielo esta estructura estable abarca grandes distancias, pero en estado líquido las agrupaciones en "estado de hielo" generalmente sólo incluyen algunas moléculas, y se están formando y desintegrando constantemente.

En una membrana de ósmosis inversa queda absorbido un pequeño volumen de agua en los espacios comprendidos entre las fibras de la membrana; allí toma la configuración parecida a la del hielo. Cuando se aplica una presión a través de la membrana, las moléculas del lado de la alta presión se incorporan a la estructura de hielo, reemplazando a las que se van "fundiendo" al otro lado. Las sustancias disueltas, salvo muy pocas excepciones, no pueden tomar esta configuración de hielo y se excluyen, por tanto. Los iones en particular están rodeados en solución por moléculas de agua colocadas de manera que apantallan su carga iónica; por consiguiente, los iones no pueden encajar en la matriz del "estado de hielo". Las pocas excepciones son las moléculas capaces de establecer un enlace de hidrógeno con el agua, tales como el metanol y la urea; es significativo el hecho que estas sustancias puedan penetrar no sólo a través de las membranas de ósmosis inversa sino también a través del hielo ordinario.

Se conoce la ósmosis inversa desde hace más de cien años, pero debido a que la permeabilidad de estas membranas tan compactas es muy pequeña, el flujo del agua a su través resulta siempre demasiado lento para ser práctico. La solución a este problema era hacer la membrana más delgada: como se señaló más arriba, el flujo del disolvente es inversamente proporcional al grosor de la membrana, pero dicho grosor no tiene influencia sobre la repulsión de los solutos. Sin embargo, no parece factible fabricar membranas lo suficiente delgadas para que den un flujo útil.

ENERGIA NECESARIA para la desalinización del agua marina; en teoría es independiente del medio de extracción, pero en la práctica las diferencias de eficacia introducen grandes diferencias en la demanda de energía. La cantidad irreducible de energía que se necesita para obtener un litro de agua pura a partir de un gran volumen de agua de mar es 0,7 kilocalorías. Si se utiliza la destilación, la energía debe aumentarse inicialmente a 540 kilocalorías; ahora bien, con un intercambiador de calor perfecto toda esta energía, excepto las 0,7 kilocalorías, podría recuperarse. De hecho gran parte se pierde. De igual forma, si se elimina la sal por congelación se necesita invertir −79 kilocalorías, y sólo parte de ellas pueden recuperarse. Debido a que la desalinización por membranas no es un proceso térmico, no necesita esta inversión de energía y la eficacia es muy elevada. El consumo de energía es de tres kilocalorías por litro (cuatro veces el mínimo teórico).

En 1960 Sidney Loeb, de la Universidad de California en Los Angeles (y ahora en la Universidad Ben Gurion de Israel), descubrió un método para fabricar membranas lo suficientemente delgadas y lo suficientemente compactas para obtener una ósmosis inversa práctica. Su método se basa en un fenómeno que se observa cuando se fabrica una membrana dejando evaporar el disolvente: después de un período inicial de secado aparece una película debido a que la solución expuesta se seca más rápidamente en la superficie que en las capas inferiores. Normalmente se evita la formación de esta película para que se forme una membrana de porosidad uniforme. Loeb dejó deliberadamente que se formara una película, después sumergió la membrana en agua para que se aglutinaran los polímeros de las capas más profundas que todavía no se habían secado completamente. El resultado fue la obtención de una membrana con una película extremadamente delgada de polímero compacto sobre una capa mucho más gruesa de material altamente poroso. Estas membranas son lo suficientemente fuertes como para soportar altas presiones, pero su película activa tiene un grosor de sólo aproximadamente un 1 por ciento del de una membrana de ultrafiltración ordinaria. Como resultado el flujo de agua es 100 veces mayor.

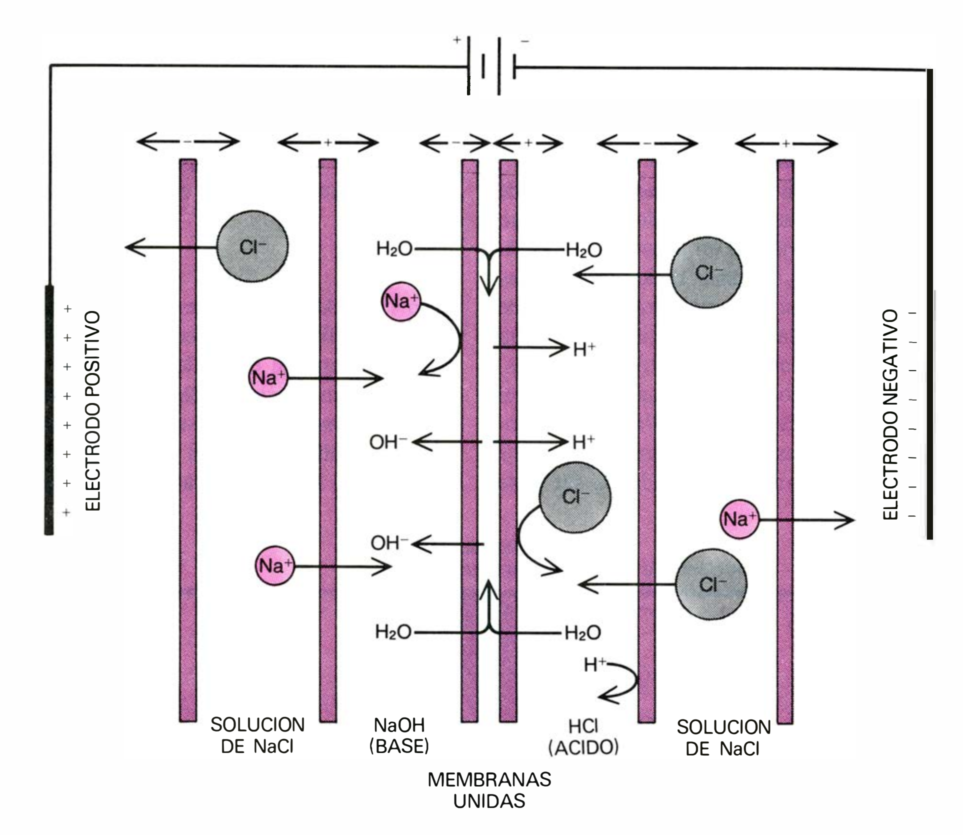
La principal aplicación de la ósmosis inversa es la extracción de agua pura de las disoluciones salinas, en particular de las aguas salobres y del agua de mar. Al contrario que las membranas de ultrafiltración, las membranas de ósmosis inversa no se hacen permeables a los iones a altas concentraciones salinas; existe, sin embargo, un límite respecto a la concentración de la solución alimentadora. Para obtener un flujo práctico de agua la presión neta a través de la membrana debe ser al menos de unos cientos de libras por pulgada cuadrada. Cuando la concentración de la solución salina es de un 3 por ciento aproximadamente, la presión osmótica llega a ser de unas 400 libras por pulgada cuadrada, y la polarización de la concentración puede aumentarla efectivamente hasta unas 600 libras por pulgada cuadrada. El coste de las bombas y el otro material limita la presión aplicada a unas 1000 libras por pulgada cuadrada, de modo que una concentración más alta es impracticable. El agua pura puede ex-

traerse del agua de mar por ósmosis inversa si puede diluirse infinitamente la sal repelida en un gran depósito (como el océano). Sin embargo, si las sales deben concentrarse para su consumo o para ser eliminadas debe combinarse la ósmosis inversa con otros procesos que no estén limitados por la concentración de la solución alimentadora.

La desalinización por ósmosis inversa da un ejemplo de una propiedad común a muchos procesos realizados con membranas: requiere menos energía que el equivalente sistema térmico de separación. Para extraer un litro de agua pura de un volumen ilimitado de agua de mar necesita un mínimo de 0,7 kilocalorías; éste es el mínimo de energía termodinámico, o teórico, y es el valor más pequeño posible, independientemente del método de separación. Sólo se puede llegar cerca del mínimo termodinámico cuando todos los pasos de un proceso son reversibles, es decir, cuando se está cerca del equilibrio. La evaporación y la condensación se pueden llevar a cabo de manera que sean casi reversibles realizando ambos procesos a la misma presión y temperatura. Sin embargo, bajo estas condiciones de equilibrio, los procesos térmicos son excesivamete lentos y requieren muchos aparatos. Puede incrementarse la velocidad operando con grandes diferencias de temperatura, pero esta desviación respecto al equilibrio y la reversibilidad representa un gran coste de energía. Por ejemplo, la extracción de un litro de agua por destilación requiere unas 540 kilocalorías, de las cuales sólo una parte pueden recuperarse en un intercambiador de calor. Por otra parte, las membranas funcionan muy bien cerca de equilibrio. Este litro de agua puede separarse por ósmosis inversa gastando unas tres kilocalorías, sólo unas cuatro veces el mínimo de energía.

La fuerza que impulsa tanto la ultrafiltración como la ósmosis inversa es la presión aplicada, pero la separación de moléculas por membranas también puede ser potenciada por otras fuerzas. En uno de estos procesos, llamado electrodiálisis, un campo eléctrico aplicado a través de una membrana o un conjunto de membranas aporta la energía. En la diálisis (y en esto se distingue de la ósmosis) son los solutos los que pasan a través de la membrana en lugar del disolvente; en la electrodiálisis los solutos son iones y las membranas son selectivamente permeables a los iones que llevan o bien cargas positivas o negativas.

En 1923 la bioquímico alemana Leonor Michaelis demostró que las membranas fabricadas a partir de nitrocelulosa secada eran selectivamente permeables a los iones positivos, pero no explicó el mecanismo de



CELULA DISOCIADORA de agua que utiliza membranas de carga fija para obtener un ácido y una base a partir de una solución salina. Las dos membranas centrales están unidas, pero existe un pequeño espacio entre ellas por el que puede entrar el agua. Como en el resto del agua, una pequeña fracción de estas moléculas (unas 10^{-7}) se ioniza espontáneamente. Después, los iones son separados por el campo eléctrico, desplazándose los iones hidrógeno (H^+) hacia la derecha y los iones hidroxilo (OH^-) hacia la izquierda. Al mismo tiempo, los iones sodio (Na^+) se concentran en el compartimiento de la izquierda de las dos membranas centrales y los iones cloruro (Cl^-) en el compartimiento de la derecha de ambas membranas. La recombinación de los iones del agua con los iones de la sal producen hidróxido sódico y acido clorhídrico.

este efecto. En 1929, Karl Sollner comenzó el primer estudio sistemático relacionando la estructura de las membranas con su funcionamiento. Junto con sus discípulos Charles Carr y uno de nosotros (Harry P. Gregor) fabricó las primeras membranas diseñadas, no sólo para obtener una gran selectividad, sino también una permeabilidad óptima.

La tecnología de la electrodiálisis está ahora muy bien desarrollada y las membranas para llevarla a cabo se fabrican en cantidades comerciales. Estas membranas tienen unos poros muy pequeños (aproximadamente un nanómetro de diámetro), y están hechas de polímeros entrelazados que tienen una gran concentración de cargas fijas. Las membranas cargadas negativamente están hechas de poliestireno con grupos sulfonato; por tanto, repelen todos los iones cargados negativamente. Son parecidas a las membranas de ultrafiltración con grupos sulfonato pero presentan una permeabilidad al agua mucho más baja y una mayor concentración de cargas fijas. Debido al gran número de grupos cargados, estas membranas son altamente selectivas incluso en soluciones salinas de hasta 175 gramos por litro.

Las correspondientes membranas cargadas positivamente están hechas de un armazón a base del polímero cloruro de polivinil bencilo, y la carga necesaria es aportada por grupos derivados del ion amonio cuaternario (NH_4^+), como el trimetilamonio. Como era de esperar, estas membranas presentan la selectividad opuesta; repelen a los iones positivos y dejan pasar los negativos. Las membranas sulfonadas, al igual que las similares empleadas en ultrafiltración, son muy resistentes a la obturación, pero las que tienen iones amonio cuaternario no lo son. Estas últimas se obturan seriamente si están expuestas a fluidos que contienen productos naturales y, por tanto, a veces es necesario realizar un pretratamiento de la solución con ultrafiltración, si se quiere hacer una electrodiálisis.

En una célula de electrodiálisis se disponen las membranas de polaridad alternante una encima de otra, y se aplica un campo eléctrico a través de todo el conjunto. Los iones positivos son atraídos por el electrodo negativo y pueden atravesar la membrana sulfonada para alcanzarlo; sin embargo, quedan frenados por la siguiente membrana, que es la que tiene iones amonio cuaternario. Los iones negativos se mueven en dirección opuesta hasta

que quedan frenados por una membrana sulfonada. El resultado es que las cavidades entre membrana se llenan alternativamente de agua purificada de solución salina concentrada.

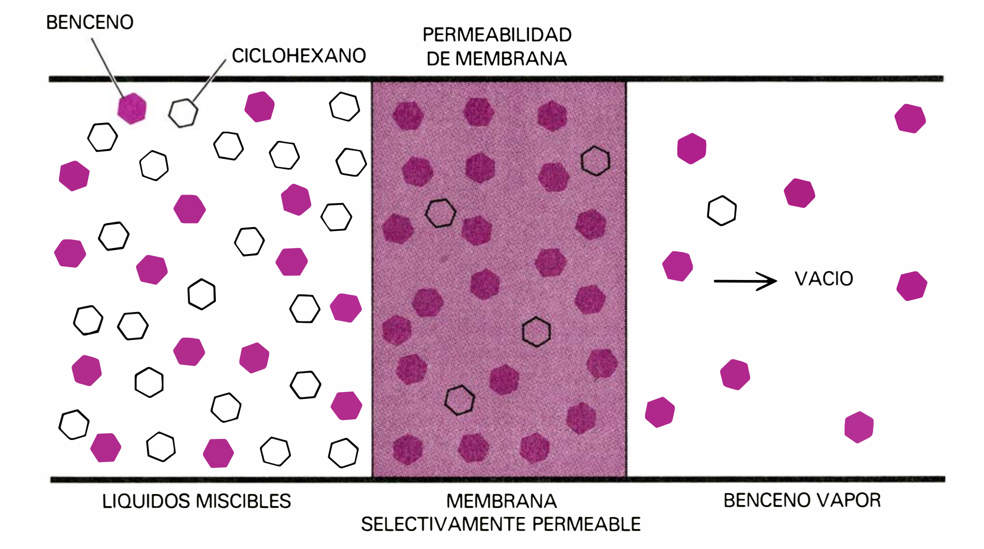
En el Japón la mayor parte de la sal de mesa se extrae del agua de mar por un método que incluye un paso de electrodiálisis. Después de haber separado los iones magnesio y sulfato por otro proceso de membranas, la electrodiálisis concentra la solución de cloruro sódico unas seis veces, hasta un 17 por ciento. El agua restante se elimina por evaporación. Para la electrodiálisis se necesita una energía de unos 300 kilovatios hora por tonelada de sal y se consumen unos 250 kilovatios hora más para evaporar el agua restante. Si todo el proceso de concentración se hiciera por evaporación, se necesitarían 2100 kilovatios hora por tonelada; así pues esta técnica híbrida consume sólo la cuarta parte de energía.

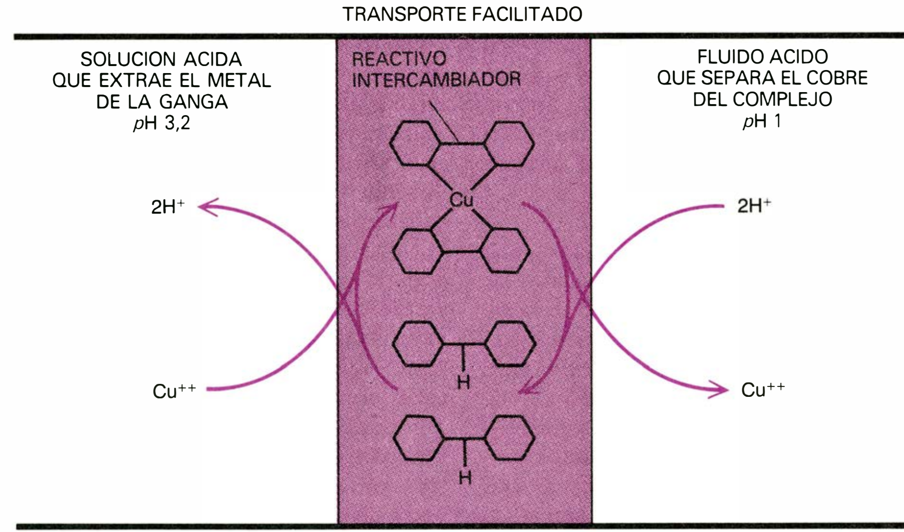
En 1913 Albrecht Bethe (fisiólogo y padre del físico Hans Bethe) montó una membrana animal entre dos soluciones e hizo pasar una corriente eléctrica a través de este sistema para observar el movimiento resultante de las moléculas de agua. Quedó sorprendido al descubrir que la solución se volvía ligeramente ácida a un lado de la membrana y ligeramente básica al otro. Estableció que este efecto era el resultado de la disociación de las moléculas de agua en iones hidrógeno (H^+) e iones hidroxilo (OH^-). El fenómeno descubierto por Bethe se consideró como una curiosidad de laboratorio hasta que se observó que las membranas de electrodiálisis comerciales también disociaban cantidades apreciables de agua. Entonces se hizo evidente que las membranas podían emplearse para obtener un ácido y una base a partir de una sal.

Es importante distinguir la disociación del agua del proceso ligeramente similar de la electrólisis; en este último los productos resultantes son el hidrógeno (H_2) y el oxígeno (O_2) moleculares. La descomposición de la molécula de agua en sus componentes atómicos exige una cantidad de energía determinada, mientras que la disociación de la molécula en iones tiene lugar espontáneamente, sin necesidad de absorber energía. En el agua pura tiene lugar en cualquier momento la ionización de una molécula de cada 10^7. La corriente eléctrica que se necesita en la técnica de disociación del agua por membrana es utilizada principalmente para separar los iones una vez formados, aunque se ha visto recientemente que el campo eléctrico aumenta ligeramente la ionización.

Se puede construir una célula disociadora de agua con membranas muy similares a las empleadas en la electrodiálisis. Se unen a lo largo dos membranas de polaridad opuesta y se colocan dos membranas más a cada lado, también de polaridad alternante. La célula completa se sumerge en una solución salina, como cloruro sódico, y se hace pasar una corriente eléctrica a través del sistema. Los iones hidrógeno e hidroxilo, que se forman en el espacio comprendido entre las dos membranas unidas, se separan según su carga; atraviesan una membrana y son repelidos por la siguiente, y así se combinan con los iones sodio y cloruro atrapados en el mismo espacio entre membranas. A un lado de la doble membrana se forma ácido clorhídrico (NCl) y al otro lado hidróxido sódico (NaOH).

Un problema que se presentó en las primeras membranas disociadoras de agua fue el resecamiento de la doble membrana al extraer del espacio interior los iones hidrógeno e hidroxilo, y las moléculas de agua que los rodeaban. A las altas densidades de corriente que se necesitan para las aplicaciones industriales, las membranas pronto se tornaban ineficaces. Recientemente uno de nosotros (Harry P. Gregor), junto con Kenneth Brennen y Bruce Benjamin, ha desarrollado métodos





EXTRACCION CON DISOLVENTES a través de una membrana, separando de esta forma las moléculas según sus distintas afinidades por un disolvente o por un transporte molecular especializado de la membrana. En una de las formas de extracción con disolventes, llamada permeabilidad de membrana (*arriba*), se coloca una mezcla de dos líquidos miscibles, como benceno y ciclohexano, a un lado de una membrana que absorba más fácilmente uno de los componentes de la mezcla que el otro. Al hacer el vacío al otro lado de la membrana, pasan a su través principalmente las moléculas del componente para el cual la membrana es más permeable, en este caso el benceno. Puede emplearse un segundo proceso llamado transporte facilitado (*abajo*) para extraer el cobre de la solución ácida que separa el metal de la ganga. La membrana que se utiliza es impermeable al cobre (y a todos los demás iones metálicos) pero contiene un reactivo intercambiador que forma de manera selectiva un complejo con el cobre. Este, se origina en donde la concentración es mayor y se difunde hacia el otro lado, donde se separa el cobre por un ácido fuerte, regenerando el reactivo.

para evitar el resecamiento uniendo cuidadosamente las membranas seleccionadas.

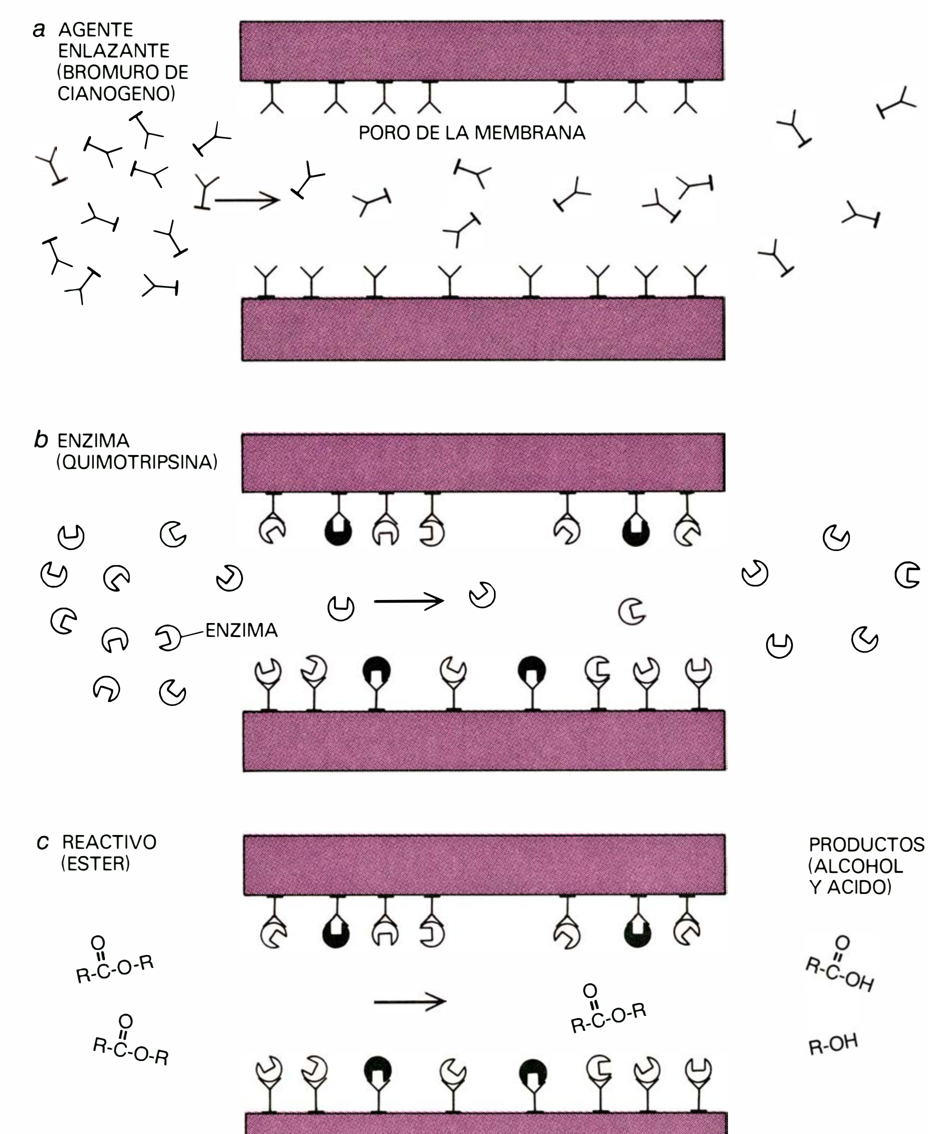
La tecnología de la disociación del agua todavía está en desarrollo, pero promete varias aplicaciones. En minería, por ejemplo, podría producirse conjuntamente el ácido y la base en el mismo lugar de trabajo. El ácido podría luego emplearse para extraer el mineral de la ganga, mientras que la base se reservaría para neutralizar los líquidos residuales antes de ser eliminados.

Una modificación de la técnica de disociación del agua permite descomponer una sal de un ácido débil y una base débil en sus gases componentes. Por ejemplo, la sal bisulfuro amónico (NH_4HS) es un producto secundario del refinado del petróleo, que se forma cuando se extrae el sulfuro del petróleo bruto por medio de un tratamiento con gas amoníaco. Una membrana disociadora de agua podría descomponer este material de desecho, produciendo amoníaco y ácido sulfhídrico.

Además del tamaño y la carga eléctrica, un tercer criterio para diferenciar moléculas es la solubilidad en varios disolventes. Estas diferencias de solubilidad se aprovechan en el proceso llamado extracción con disolventes. En su forma más simple, la extracción con disolventes consiste en purificar una sustancia disuelta en agua agitando la solución en otro disolvente, como el petróleo, que no es miscible con el agua pero en el que la sustancia es más soluble.

La extracción con disolventes a través de membranas se conoce por lo menos desde 1913, cuando Fritz Haber y Reinhardt Beutner demostraron que una fina película de un derivado del petróleo se podía emplear como membrana en dos tipos de procesos de extracción. En uno de ellos, denominado ahora permeabilidad de membrana, se coloca una mezcla de dos líquidos miscibles, como el benceno y el ciclohexano por ejemplo, a un lado de la membrana y se hace el vacío al otro lado. Si la membrana es más permeable a uno de los dos componentes, verbigracia al benceno, el vapor que aparece en el lado que se ha hecho el vacío es en gran parte benceno. Al contrario que la destilación, en el proceso de permeabilidad de membrana sólo se necesita que se vaporice uno de los componentes de una mezcla, y de aquí que se reduzca la energía requerida.

Otra forma de extracción con disolventes ideada por Haber y Beutner, denominada transporte facilitado, se asemeja en algunos aspectos al mecanismo que se supone que debe funcionar para el transporte activo de las moléculas a través de las membranas biológicas. Se colocan disoluciones acuosas, una de ellas conteniendo la sustancia que se quiere aislar, a ambos lados de una membrana impregnada de un derivado del petróleo. Entonces una molécula móvil de la membrana forma un complejo con la sustancia seleccionada; el complejo se difunde a través de la membrana y se disocia en el lado opuesto.

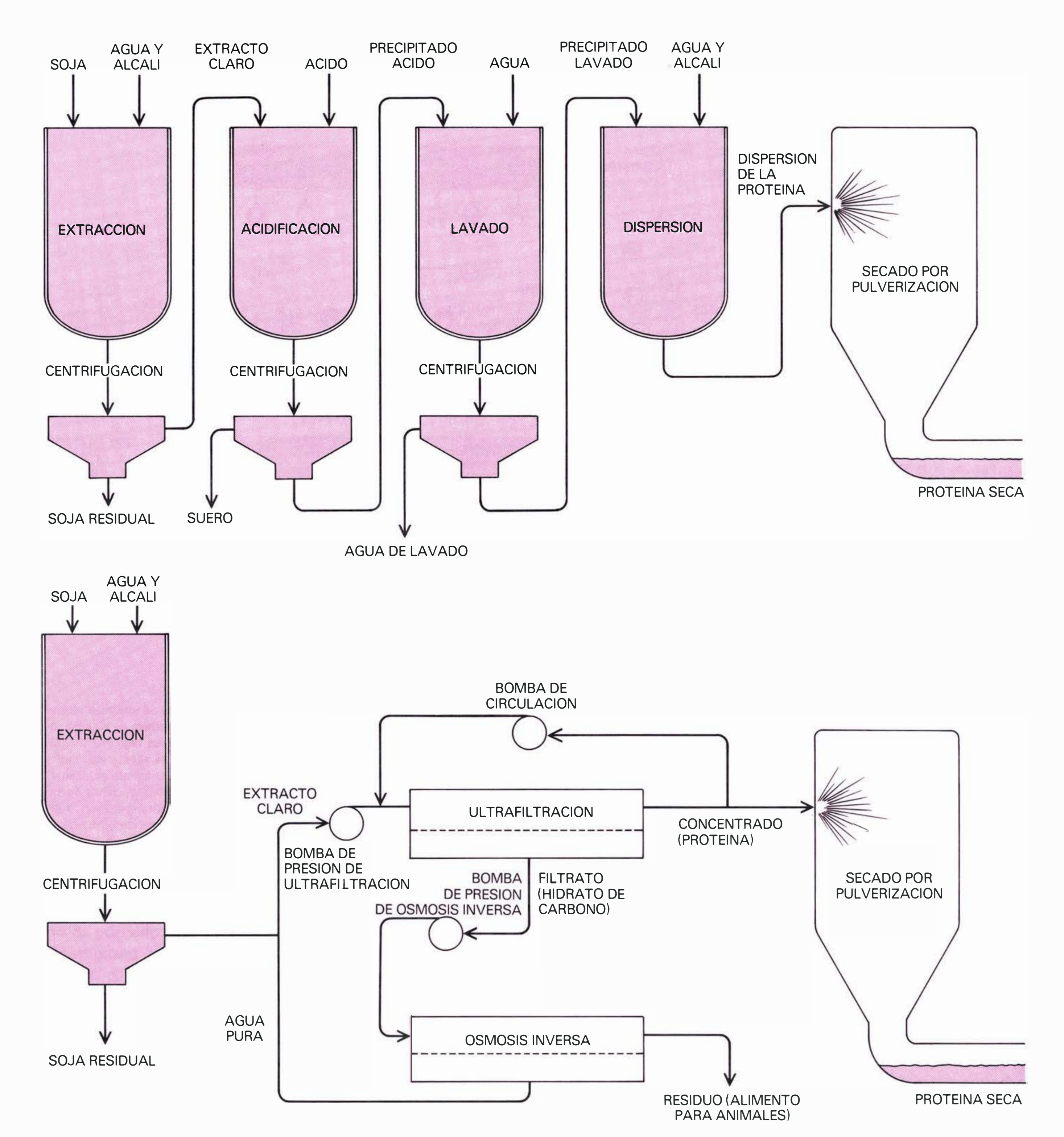


REACTOR ENZIMATICO: puede construirse introduciendo un enzima en los poros de una membrana de polímeros. Primero se hace pasar a través de la membrana un agente enlazante, el bromuro de cianógeno (CNBr) por ejemplo, que se une a las paredes de los poros (*a*). Luego se introduce, también bajo presión, el enzima propiamente dicho, en este caso el enzima digestivo quimotripsina, que se une al agente enlazante (*b*). Algunas moléculas del enzima se desactivan, principalmente aquellas que se unen al agente enlazante por un punto cercano al centro activo del enzima, pero puede conservarse hasta un 80 por ciento de la actividad del enzima. Una vez preparada la membrana se hace pasar a su través un sustrato del enzima, y al otro lado aparecen los productos de la reacción enzimática (*c*). Por ejemplo, el enzima digestivo quimotripsina, que vemos se une al enlace enlazante convierte un éster en un ácido orgánico y un alcohol.

Lawrence Brandlein y uno de nosotros (Harry P. Gregor) demostraron que por este método se puede extraer el cobre de la solución en la que se encuentra una vez separado de la ganga. El cobre está presente junto con otros metales en una solución débilmente ácida, que se coloca a un lado de una membrana impregnada de una solución de queroseno. Ninguno de los iones metálicos es soluble en queroseno y, por tanto, la membrana es impermeable a ellos. Los iones cobre son absorbidos por la membrana por medio de un reactivo orgánico que forma, de manera selectiva, un complejo soluble en queroseno con el cobre, pero no con los otros metales. Al otro lado de la membrana se separa el cobre del complejo mediante un ácido fuerte. De esta manera con una sola operación se extrae el cobre de la solución en la que se encuentra y se aísla de los metales contaminantes. La extracción con disolventes a través de membranas también puede aplicarse al reciclaje de los combustibles nucleares.

En otro proceso de membranas no se

explotan las capacidades selectivas de la membrana; la membrana se utiliza principalmente como matriz inerte a la que se adhieren los enzimas. En 1954 Nikolaus Grubhofer demostró que un enzima se podía unir químicamente a una matriz sintética sin pérdida de actividad. Los trabajos subsiguientes de Ephraim Katchalsky y Georg Manecke llevaron al desarrollo de los reactores químicos en los que los enzimas están unidos a pequeñas esferas porosas e insolubles. (Katchalsky, que más tarde tomó el nombre de Katzir, acaba de abandonar la presidencia del estado de Israel.) Estos reactores funcionan ahora comercialmente (por ejemplo, convirtiendo la glucosa en azúcar invertido), pero tienen ciertas deficiencias. Para enlazar el enzima a las esferas se necesita normalmente la presencia simultánea del enzima y de un agente enlazante corrosivo, con lo que



EXTRACCION DE PROTEINA de la soja; podría simplificarse con el uso de membranas sintéticas. En la extracción convencional (*arriba*) la proteína primero se disuelve en álcali, junto con otras sustancias, entre las que hay ciertos hidratos de carbono indeseables. Después, la proteína se precipita con ácido, se lava con agua para eliminar la mayor parte de las impurezas, se dispersa de nuevo en álcali y finalmente se seca por pulverización. En cada uno de estos pasos los residuos en suspensión se eliminan por centrifugación. El proceso que utiliza membranas empieza también con una extracción con álcali, pero las sustancias indeseables se eliminan por ultrafiltración. Entonces el líquido filtrado puede someterse a una ósmosis inversa para concentrar el residuo, que está formado principalmente por hidratos de carbono, y recuperar el agua pura. El tratamiento con membranas exige menos energía y deja menos residuos.

gran parte del enzima a menudo se destruye. Además, debido a que las sustancias reaccionantes y los productos de la reacción circulan solamente por difusión, el proceso es más bien lento.

Paul Rauf y uno de nosotros (Harry P. Gregor) han demostrado que estas dificultades pueden evitarse enlazando el enzima a una membrana cuyos poros sean unas cinco veces mayores que las moléculas del enzima. La membrana se prepara bombeando a través de los poros un agente enlazante, como el bromuro de cianógeno (CNBr), que forma un compuesto intermedio inestable, capaz de unirse al enzima. Después de eliminar el exceso de agente enlazante se introduce el enzima bajo presión. Controlando la presión y el tiempo de reacción, los poros pueden recubrirse de una densa capa de enzima, que mantiene hasta el 80 por ciento de su actividad inicial. El enzima está unido a la membrana y la actividad enzimática de una molécula determinada sólo se verá afectada si la unión tiene lugar cerca del centro activo.

En un experimento se impregnó la membrana del enzima quimotripsina, que normalmente digiere las proteínas pero que también es capaz de descomponer un éster en un ácido orgánico y un alcohol. Se encontró que la velocidad de transformación de los ésteres aumentaba con la presión de la corriente alimentadora, y sólo a una presión aplicada de unas cien libras por pulgada cuadrada se alcanzaba la capacidad catalítica del enzima.

Se han descrito más arriba varias aplicaciones de la tecnología de las membranas sintéticas. Entre otras aplicaciones potenciales pueden servir de ejemplos ilustrativos tres industrias que pueden utilizar membranas a gran escala. Son el tratamiento de las aguas residuales, la industria de transformación de alimentos y el propuesto cultivo de plantas verdes como fuente de combustible.

La depuración convencional de las aguas de alcantarilla utiliza una combinación de grandes tanques de sedimentación, cultivos bacterianos y mecanismos de concentración de los residuos para depurar las aguas residuales y concentrar los restos sólidos. El primer tratamiento, mediante el cual se dejan sedimentar las partículas mayores, probablemente no puede mejorarse por medio de la tecnología de membranas. Sin embargo, el segundo tratamiento que se utiliza hoy día podría sustituirse por una ultrafiltración con membranas no obturables. Kang Hsu, Sadi Mizrahi y uno de nosotros (Harry P. Gregor) han demostrado que el líquido resultante de esta filtración es de una calidad igual o superior al que resulta del segundo tratamiento convencional. En realidad, ya que las bacterias y virus no pueden atravesar la membrana, el líquido obtenido es casi estéril. Y lo que es más importante, ya que los sólidos de la corriente que se filtra no se descomponen, se depositan y se concentran más rápidamente, facilitando su eliminación final. Los residuos concentrados pueden mantener su propia combustión en un horno crematorio; además, por contener de un 20 a un 30 por ciento de proteína, constituyen una fuente potencial de fertilizante nitrogenado.

Ya que las unidades de ultrafiltración

CULTIVO DE ALGAS COMO COMBUSTIBLE; se trata de una tecnología a considerar que utilizaría varios tratamientos con membranas. Las algas podrían ser recolectadas por medio de una membrana de ultrafiltración con poros relativamente grandes, introduciendo después la suspensión concentrada de células de algas en un digestor bacteriano donde se produciría de hecho el combustible (como metano). Los productos secundarios de la digestión se eliminarían por medio de otra membrana de ultrafiltración y se podrían descomponer en una célula disociadora de agua. El amoniaco (NH_3) y el dióxido de carbono (CO_2) son sustancias nutritivas de cierto valor y podrían ser devueltas al estanque o tanque donde estén las algas. Los ácidos orgánicos, que tienen valor comercial, podrían separarse por medio de una membrana de extracción con disolventes. Los ácidos inorgánicos y las bases podrían concentrarse como sales por electrodiálisis.

son modulares, la instalación de pequeñas plantas de tratamiento podría ser económicamente factible. De hecho, se está considerando la posibilidad de emplear la ultrafiltración junto a las fosas sépticas en las casas particulares. La ultrafiltración también podría tener especial interés para el tratamiento de las aguas residuales de las regiones desérticas, de las que deben eliminarse las sales junto con otros contaminantes.

En la elaboración y refinado de muchos alimentos, como el queso y la soja, se obtienen líquidos residuales que contienen grandes cantidades de sustancias nutritivas, pero en concentraciones demasiado bajas para que su recuperación sea económicamente factible. Pero si tales líquidos se vierten sin más, pueden ocasionar una grave contaminación.

En la elaboración del requesón, el líquido residual es el suero de la leche, que se produce en los Estados Unidos a razón de cinco mil millones de galones por año. Este suero contiene aproximadamente un 3 por ciento del azúcar lactosa, un 1 por ciento de la proteína de la leche lactoalbúmina, cantidades menores de residuos sólidos y ácido láctico. Podría concentrarse por evaporación, secarse pulverizándola y ser utilizado como alimento para los animales, pero la energía que se necesita para este proceso hace que sea poco atractivo, económicamente. Sin embargo, se ha demostrado por estudios recientes que de la ultrafiltración del suero de la leche se obtiene un concentrado que contiene de un 10 a un 20 por ciento de proteína, junto con un líquido claro que contiene lactosa, sales y ácido láctico. La lactosa puede concentrarse posteriormente por ósmosis inversa hasta un 24 por ciento, y entonces el azúcar puro puede recuperarse por evaporación. O bien, como alternativa, la solución de lactosa puede hacerse pasar a través de un reactor enzimático que lo convierte en los azúcares más simples glucosa y galactosa, que son más apetecibles que la lactosa.

La tecnología de membranas puede emplearse de manera similar en la purificación de otros residuos y productos secundarios del refinado de los alimentos. Por ejemplo, la extracción de proteína de la harina de soja produce líquidos que contienen, además de las proteínas deseadas, hidratos de carbono y sales; éstos podrían separarse y purificarse por una combinación de ultrafiltración y ósmosis inversa. La molienda del maíz produce líquidos residuales que se denominan aguas muy solubles; una planta típica debe desechar 20 millones de galones al día, que pueden evaporarse, al precio de 200.000 dólares al día, hasta un 50 por ciento y mezclarlos con el forraje como alimento para los animales. La ultrafiltración podría separar toda la materia suspendida y la proteína disuelta y recuperar el 99 por ciento del agua, la cual podría después purificarse por ósmosis inversa hasta alcanzar la calidad de agua potable. El residuo de este segundo tratamiento por membranas, que contiene el 13 por ciento de materia sólida, es el único que se necesita concentrar posteriormente por evaporación a fin de poderse utilizar como alimento para los animales. Una evaluación económica de los procesos realizados a través de membranas sugiere que sólo se necesitaría el 20 por ciento de la energía empleada en la evaporación directa del líquido sobrante.

Un método que puede tener interés para la obtención de energía solar estriba en el cultivo de cosechas que puedan convertirse en líquidos combustibles, como el metanol. Entre las plantas que pueden considerarse como cosechas combustibles están las algas, que transforman eficazmente la energía solar y que pueden crecer en una gran variedad de ambientes. Sin embargo, existen dos impedimentos importantes para esta economía de energía: las células de las algas, que tienen un tamaño comprendido entre uno y 10 micrometros, se tendrían que recolectar de manera rápida y barata de los estanques en los que crecieran, y se tendría que mantener el enorme suministro de sustancias nutritivas que requieren estas plantas en su crecimiento. La tecnología de membranas podría aliviar tales problemas.

Las células de algas están presentes en el agua a una concentración de sólo 0,1 o 0,2 gramos (peso seco) por litro, y han de poderse concentrar al menos hasta 10 gramos por litro antes de introducirlas en un digestor bacteriano donde se formarían los combustibles. Las algas pueden concentrarse por precipitación química o por centrifugación, pero los costes de estos métodos resultan prohibitivos para una cosecha que en su mayor parte ha de quemarse y que, por tanto, debería resultar barata. La ultrafiltración ofrece una alternativa interesante; y ya que las células son más bien grandes comparadas con el orden del diámetro de los poros, puede emplearse una membrana poco compacta con un flujo elevado. En los experimentos realizados por Joaquín Errazuriz y uno de nosotros (Harry P. Gregor) se elevó la concentración del alga *Selenestrum* de 0,2 gramos por litro a 17 gramos por litro por ultrafiltración. Las células no se rompen y el líquido resultante es completamente claro. Se ha estimado que en los cultivos de agua dulce y de agua marina a gran escala los costes por ultrafiltración no llegarían a los tres centavos de dólar

por 1000 galones, lo cual cae dentro de los límites de la factibilidad económica.

El suministro de sustancias nutritivas para las algas es un problema más difícil. John H. Ryther y Joel Goldman, de la Woods Hole Oceanographic Institution, han estimado que si se dedicara toda la producción de la industria de los fertilizantes de Estados Unidos a estas cosechas combustibles, solamente se aportaría de un 1 a un 3 por ciento de la cantidad necesaria. Esta demanda sólo puede reducirse conservando las sustancias nutritivas, proceso que puede inducir varias tecnologías de membrana.

Actualmente está funcionando un prototipo de sistema para obtener energía de las algas, construido por Daniel Omstead, Thomas Jeffries y uno de nosotros (Harry P. Gregor). El material del digestor bacteriano es ultrafiltrado para eliminar los productos secundarios, y se beneficia de una digestión considerablemente acelerada. Se ha diseñado la digestión para que produzca metano; los productos secundarios son bicarbonato amónico, sales de ácidos orgánicos y sales inorgánicas comunes. El líquido que contiene estas sustancias se hace pasar a través de una célula disociadora de agua y se obtiene amoníaco de dióxido de carbono. El dióxido de carbono se disuelve en una solución básica y se devuelve al digestor para aumentar la producción de metano; el amoníaco puede reciclarse o puede venderse como fertilizante. Los ácidos orgánicos que tienen valor comercial son separados por extracción con disolventes. Finalmente, se utiliza la ósmosis inversa y la electrodiálisis para purificar las sustancias nutritivas que quedan, y así poderlas reciclar, al objeto de concentrar los residuos sólidos que han de eliminarse y recuperar el agua pura. Naturalmente, incluso en este elaborado sistema de ciclos cerrados, se pierden algunas sustancias nutritivas. Un posible modo de compensar dicho déficit sería empleando los residuos de las aguas de albañal como fertilizante. Cuando se han concentrado los residuos por ultrafiltración, pueden transportarse sin grandes dispendios económicos.

Merece la pena tener en cuenta que todas las tecnologías de membrana tienen sus modelos en la célula viva. En realidad, un solo órgano, el riñón, lleva a cabo actividades comparables a las de todas las seis tecnologías de membrana que se han expuesto aquí: la ultrafiltración, la ósmosis inversa, la electrodiálisis, la disociación del agua, la extracción con disolventes y la promoción de reacciones químicas por medio de enzimas unidos a la membrana. Las membranas sintéticas no son todavía comparables a las biológicas en velocidad, solidez, selectividad y eficacia.

El kiwi

Nueva Zelanda, donde esta ave áptera vive hoy en día, careció de mamíferos durante 80 millones de años. Al llenar los nichos ecológicos que hubieran sido ocupados por los mamíferos, los kiwis desarrollaron caracteres de mamífero

William A. Calder III

El reino animal es tan rico en su diversidad que la evolución parece haber sido un proceso al azar. Considérese el kiwi, esa famosa ave de Nueva Zelanda que no puede volar. La selección natural es en parte un proceso de probabilidad, afectado por mutaciones al azar y accidentes históricos de ascendencia y localización; al mismo tiempo opera bajo determinadas compulsiones. Por ejemplo, las leyes físicas determinan qué tamaño puede alcanzar un ave y ser todavía capaz de volar. Durante su corta vida no es probable que ningún humano pueda ser testigo de la interacción entre ambiente y herencia, que da origen a un organismo adaptado a un nicho ambiental disponible, pero la historia natural del aislamiento biológico presenta algunos casos sugestivos. El del kiwi se cuenta entre los más fascinantes.

Con frecuencia se oye que las aves presentan adaptaciones evolutivas maravillosas. Los ejemplos que se aducen pueden ser la larga migración del charrán ártico, con sus notables logros de navegación; la envergadura alar de de esos maestros del vuelo cernido, los grandes albatros; la coevolución de los diminutos colibríes y de las plantas que poseen flores tubulares rojas y las complicadas exhibiciones nupciales de los pergoleros. ¿Puede decirse algo parecido de un ave que ha perdido su capacidad de volar, que hoza en busca de lombrices, que excava en el excremento para poner un ridículo huevo enorme y que es conocido por el lego sobre todo a través de la etiqueta de una marca de crema para el calzado? Mi respuesta, dado el contexto de aislamiento y de linaje, es un sí rotundo.

Hay tres especies de kiwi. El kiwi común o pardo *(Apteryx australis)* se encuentra en las islas Norte y Sur de Nueva Zelanda y también en la isla Stewart, frente a la punta más austral de la isla Sur. El kiwi manchado menor *(A. oweni)* se encontraba hasta hace poco tanto en la isla Norte como en la isla Sur, pero hoy día se halla sólo en esta última. El kiwi manchado grande *(A. haasti)* se encuentra asimismo sólo en la isla Sur. Estas aves no se conocen de ninguna otra parte del mundo. Los kiwis son los miembros más pequeños de las siete familias de aves ratites, es decir, aves no voladoras que carecen del esternón carenado al que se fijan los grandes músculos pectorales que accionan las alas en las aves voladoras. (Ratite procede del latín *ratis*, que significa balsa, un navío sin quilla.) Entre las demás ratites se cuentan el avestruz, el emú y el gigantesco moa, hoy extinto, de Nueva Zelanda. Además de carecer de esternón carenado, el kiwi comparte con las demás ratites un plumaje adulto que tiene una estructura similar al plumón juvenil de otras aves. Las abundantes plumas del kiwi son largas y flexibles; carecen del mecanismo de trabado usual, lo que da al plumaje una apariencia hirsuta, como de pelaje. Las plumas no se desordenan cuando el kiwi excava las madrigueras en las que pasa las horas diurnas, y proporcionan aislamiento en su hábitat subterráneo, frío y húmedo, y en ambientes que van desde un clima semitropical, en la parte más septentrional de la isla Norte, hasta un clima virtualmente subpolar, en los Alpes Australes de la isla Sur.

El kiwi no posee cola externa, y sus alas se hallan reducidas a vestigios que se esconden bajo el plumaje. Su sentido del tacto está aumentado por medio de cerdas rictales bien desarrolladas en la base del pico y por una región sensitiva en su punta. Sus patas son muy fuertes; sirven al ave no sólo para la locomoción, sino también para la excavación y la lucha. La postura del kiwi mientras se halla excavando recuerda la actitud de un pato nadador.

El kiwi se nutre de insectos, bayas, semillas y otros materiales vegetales, pero su principal alimento son las lombrices de tierra. Nueva Zelanda tiene una fauna de lombrices de tierra particularmente rica, y el kiwi tiene a su disposición 178 especies nativas y 14 especies introducidas de lombrices. La técnica de búsqueda del ave consiste en ir hozando el suelo con su largo pico y oliendo las lombrices. Betsy G. Bang, de la Universidad Johns Hopkins, encontró que, en relación con el tamaño del prosencéfalo, el bulbo olfativo del kiwi es el segundo en tamaño de todas las aves que ella estudió. En el kiwi la relación del diámetro del bulbo olfativo al diámetro del hemisferio cerebral es del 34 por ciento. (El petrel de las nieves presenta la mayor relación: 37 por ciento.) Con relación a la función, esta prueba es sólo circunstancial, pero se ve apoyada en el resultado de experimentos realizados por Bernice M. Wenzel, de la Universidad de California en Los Angeles. En estos experimentos se ofrecía a los kiwis macetas llenas de tierra, en cuyo fondo había recipientes con comida o bien con sólo tierra; en este último caso, los kiwis hacían caso omiso de la maceta y no hurgaban en ella.

La biología del kiwi puede apreciarse mejor en un contexto de aspectos aparentemente dispares: el ambiente natural de Nueva Zelanda, la evolución conjunta de las ratites y la función de los huevos de las aves y de la cáscara de los mismos. La topografía de Nueva Zelanda es extraordinariamente escarpada, elevada por la acción de fallas y por vulcanismo y esculpida por el hielo glaciar. Las precipitaciones suelen ser abundantes, y antes de que los colonos europeos empezaran a alterar el paisaje éste era predominantemente forestal. En las elevaciones y en las latitudes más bajas los bosques constituían una mezcla de árboles de hoja ancha y de coníferas podocarpáceas, un grupo de árboles propio del Pacífico sudoccidental y de Chile. Por encima de este cinturón de bosque mixto y hasta el límite de la vegetación arbórea, de los 900 a los 1500 metros, el bosque era principalmente de haya austral *(Nothofagus)*. También existían regiones de pradera natural, sobre todo en la zona de sotavento, más seca, de las principales cordilleras, como el área al este de los Alpes Australes.

Con excepción de dos especies de murciélagos, no hubo mamíferos en Nueva Zelanda hasta que el hombre llegó a las

islas en el primer milenio a. de C. De ahí que la evolución de las dos familias de ratites (los kiwis y los moas) pudiera realizarse libre de la depredación y de la competencia que eran la regla dondequiera que existían mamíferos terrestres. Aquí tuvieron las ratites su mayor éxito en cuanto a diversificación. W. R. B. Oliver, del Canterbury Museum, en Christchurch, un estudioso de los recientemente extinguidos moas asignó sus restos semifosilizados a 22 especies distintas. A partir de otro análisis más reciente, Joel Cracraft, del Medical Center de la Universidad de Illinois, concluye que los fósiles representan sólo 13 especies. No obstante, unidas a las tres especies actuales de kiwi, ello constituye un total de 16 especies de ratites que se hallan distribuidas sobre una superficie cuyo tamaño es equivalente al del estado de Colorado, aunque se extiende sobre una diversidad de latitudes igual a la que existe entre Daytona Beach (Florida) y Goose Bay (Terranova).

Algunos de los moas que evolucionaron en el aislamiento insular tenían sólo medio metro de altura cuando adultos; otros alcanzaban una altura de tres metros. Dean Amadon, del American Museum of Natural History, estima que las dos razas de la especie mayor, *Dinornis giganteus*, pesaban entre 230 y 240 kilogramos, más del doble del peso de un avestruz. En contraste, un kiwi pardo adulto no alcanza un tamaño superior al de una gallina doméstica y pesa por término medio 2,2 kilogramos.

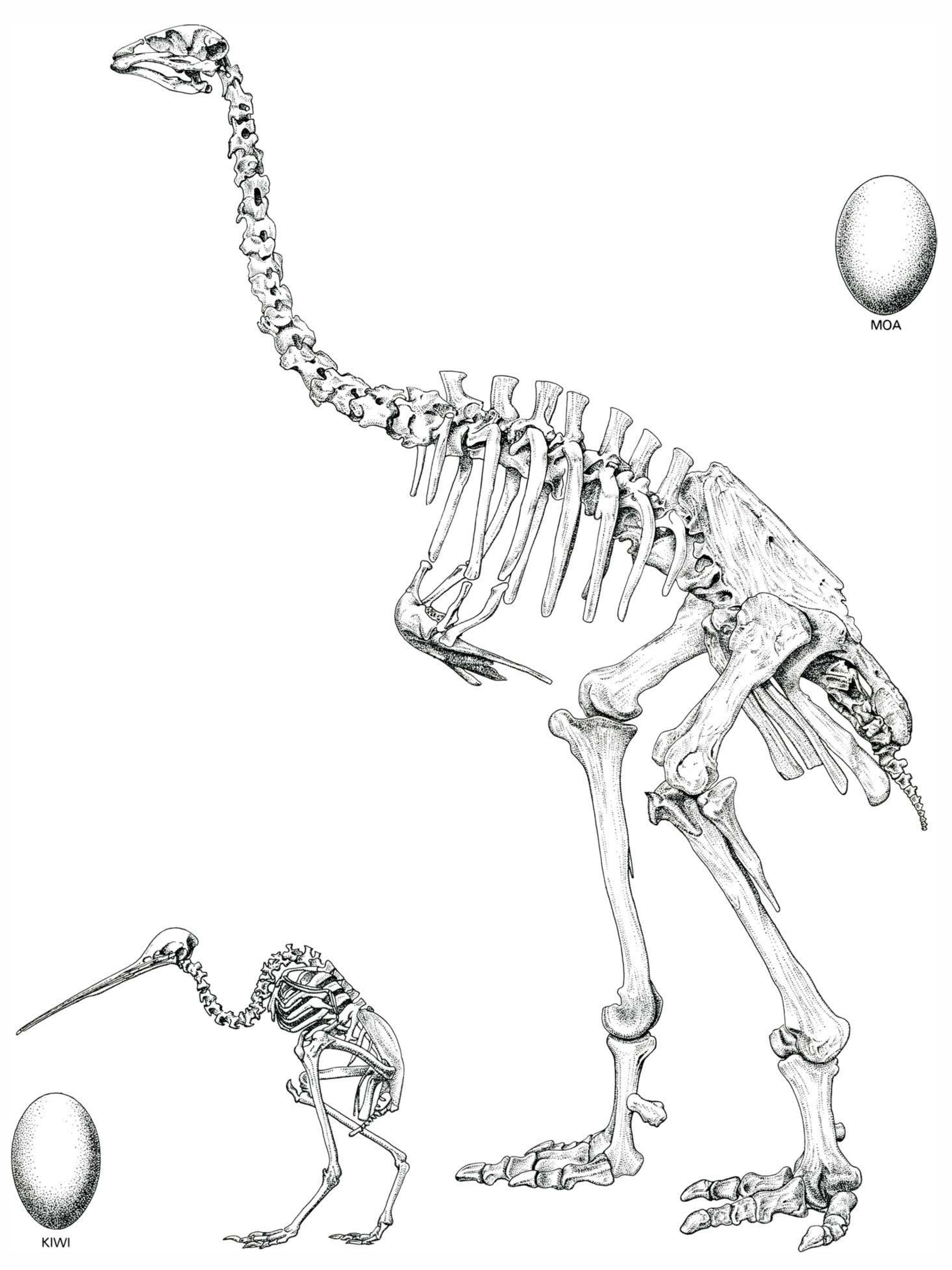
Aparte de los murciélagos, los primeros mamíferos que invadieron Nueva Zelanda fueron hombres: un grupo de polinesios navegantes, conocidos hoy día únicamente como los cazadores de moas, que llegaron a las islas en el siglo IX a. de C. No se sabe a ciencia cierta si llevaron consigo perros y ratas; si no lo hicieron, éstos debieron llegar con incursiones polinesias posteriores.

Al tratarse de los mayores animales terrestres en Nueva Zelanda hasta la llegada del hombre, los moas constituyeron una presa evidente para los invasores. Aunque algunas especies de moas pudieron haber desaparecido incluso antes de la llegada de los primeros seres humanos, los que sobrevivieron fueron probablemente aniquilados antes de la época del primer contacto con los europeos: la visita del capitan Cook en 1769. La principal invasión de mamíferos empezó a principios del siglo XIX, cuando los colonos europeos introdujeron 22 tipos distintos de animales, entre los que se contaban cerdos, ganado ovino y vacuno. Se introdujeron de intención muchos inmigrantes exóticos como animales de caza, o bien escaparon de la domesticidad, y empezaron a causar estragos ecológicos.

Desde el momento de su llegada los europeos ejercieron presiones ambientales sobre la flora y la fauna de las islas, que habían permanecido aisladas durante mucho tiempo, de distintas maneras. El hábitat forestal del kiwi se redujo debido a un asalto doble: la tala de los árboles nativos y el desbroce de tierra para el pastoreo de las ovejas. En años recientes muchos bosques han sido desbrozados y replantados con pino de Monterrey importado, pues este árbol crece más rápidamente que las especies nativas. Se han multiplicado las trampas y los cebos envenenados en un intento de controlar a los mamíferos importados que se han convertido en plagas. Por lo que se sabe actualmente de la persistencia de los venenos en las cadenas alimentarias de otros lugares del mundo, parece probable que el veneno haya afectado de manera importante a los animales nativos, kiwi incluido. De hecho, F. C. Kinsky, del Dominion Museum, en Wellington, ha citado casos de kiwis muertos por envenenamiento. El kiwi está amenazado por el trampeo, los automóviles, la depredación por parte de perros y el aclarado de terreno mediante quemas.



EL KIWI PARDO O COMUN *(Apteryx australis)* **es la principal especie de las tres todavía presentes en Nueva Zelanda. Tiene un aspecto hirsuto porque su plumaje se parece al plumón juvenil de otras aves. Se alimenta principalmente buscando lombrices con su largo pico; come también insectos, bayas y semillas.**



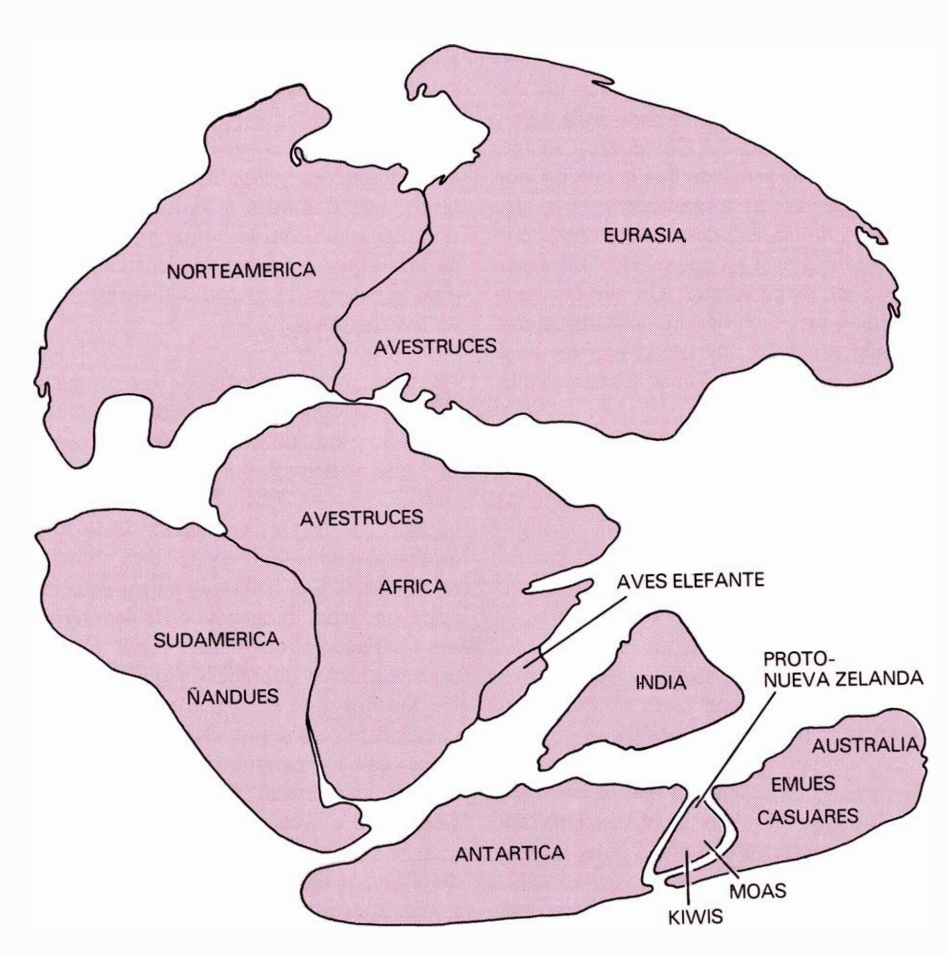
ESQUELETOS DEL KIWI Y DEL MOA, un pariente gigante extinguido del kiwi, dibujados aquí de perfil. Los ejemplares tienen 30 cm y 90 cm de altura al hombro, respectivamente; de hecho, la especie de moa representada aquí es una de las más pequeñas. Ambos miembros de la subclase de aves denominada ratites, o sin quilla, el kiwi y el moa son de anatomía parecida, con un esternón plano, alas vestigiales y patas fuertes. Sus huevos tienen aproximadamente la misma longitud; sin embargo, el huevo del moa posee un diámetro ligeramente mayor y así debió pesar alrededor de un 35 por ciento más que el del kiwi.

Actualmente los kiwis están protegidos, pero nadie sabe con certeza por qué está desapareciendo el kiwi manchado menor. Brian Reid y Gordon Williams, del New Zealand Wildlife Service, creen que el ave desapareció de la isla Norte antes de la llegada de los colonos europeos. No parece posible que fuera cazado hasta su extinción por los polinesios, puesto que el kiwi pardo, que es igualmente vulnerable a la caza, todavía está presente en la isla Norte. Para complicar el rompecabezas, el proceso parece que es continuo. Reid señala que aunque el kiwi manchado menor no ha sido eliminado en la isla Sur, raramente se le ve allí. Los moas y los kiwis, estrechamente emparentados, tienen por todo el mundo parientes más lejanos. Además del avestruz de Africa y del emú de Australia, las ratites incluyen los casuares de Australia y Nueva Guinea, los ñandúes de Sudamérica y las extinguidas aves elefante de Madagascar y las islas Canarias. ¿Cómo evolucionaron estas aves carentes de la capacidad de volar?

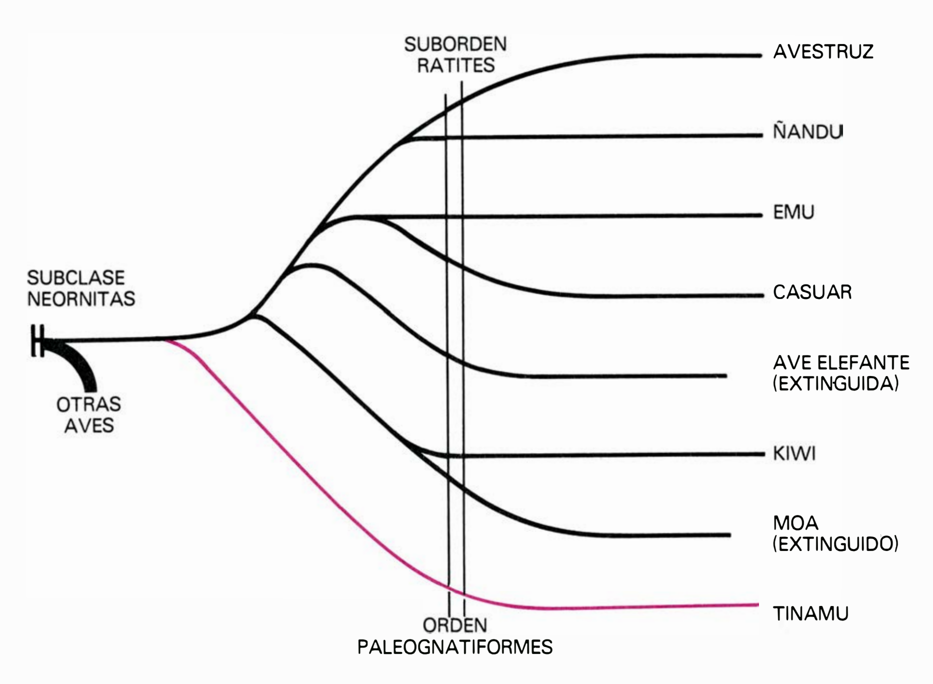
La anatomía de las ratites demuestra claramente que sus antepasados eran aves voladoras. Por ejemplo, los huesos de la muñeca de las ratites están fusionados para ofrecer más resistencia en el punto en el que las plumas remeras primarias deberían insertarse, aunque sus alas están reducidas en tamaño y función. El vuelo requiere una gran demanda de oxígeno, un soporte mecánico poderoso y una coordinación precisa, y la fisiología de las ratites muestra todavía que sus antepasados cumplían estos requisitos. Estas aves han conservado el sistema de sacos aéreos para la eficaz ventilación de los pulmones, la fusión de las vértebras inferiores con la pelvis y el acortamiento de la vértebra caudal para aumentar la resistencia y un gran cerebelo para la coordinación.

Se cree que todas las aves tuvieron un antepasado común, del tamaño de una paloma, que surgió en algún momento del Jurásico, hace 150 millones de años, de modo que es razonable preguntarse qué factores concretos favorecieron la selección natural de aves tan grandes como las ratites. Las ventajas de un cuerpo grande comprenden una velocidad creciente a la carrera, un dominio vital mayor y la posibilidad de ser dominante sobre otros animales. Otra ventaja es una resistencia acrecentada en los momentos en los que hay poco alimento; ello se debe a que la cantidad de energía en forma de alimento que un animal puede almacenar aumenta en proporción lineal a su masa, mientras que la tasa metabólica a la que se quema la energía almacenada no aumenta.

Al mismo tiempo, un cuerpo de mayor tamaño supone unas necesidades alimen-



LA DISTRIBUCION ACTUAL DE LAS RATITES puede explicarse como el resultado de la pérdida de la capacidad de volar de un antepasado común en una época anterior a aquélla en que la deriva continental dividió el antiguo continente de Gondwana en los continentes australes de la actualidad. Según Joel Cracraft, del Medical Center de la Universidad de Illinois, los kiwis y los moas son anatómicamente más primitivos que las demás ratites. Su aislamiento puede atribuirse a la separación de Nueva Zelanda de Antártica occidental hace 80 millones de años, como se indica aquí. Las aves elefante de Madagascar surgieron de la estirpe ancestral que permaneció en el continente de Gondwana, como los casuares y emúes de Australia y Nueva Guinea, los avestruces de Africa y Eurasia y los ñandúes de Sudamérica.



ARBOL FILOGENETICO DE LAS RATITES, basado en estudios anatómicos realizados por Cracraft. Este coloca las siete familias en un tronco común con los tinamúes, aves sudamericanas que parecen perdices y que han conservado la capacidad de volar *(color).* Cracraft considera que los avestruces y los ñandúes son las ratites más especializadas y que están estrechamente emparentadas. Los estudios inmunológicos sugieren, en cambio, que los ñandúes están más relacionados con los kiwis de lo que se da a entender aquí.

tarias totales superiores, lo que es potencialmente desventajoso. Quizás una desventaja incluso más grave es que un ave grande puede pesar demasiado para volar. C. J. Pennycuick, del University College, de Nairobi, ha señalado que la energía que necesita un ave para mantenerse en el aire aumenta de manera desproporcionada con respecto a la energía que pueden dispensar los músculos pectorales. Un ave que pese más de unos 12 kilogramos no puede emprender el vuelo. De modo que las aves gigantes han hecho un mal negocio evolutivo.

Otro rompecabezas evolutivo resultaba mucho más confuso hasta hace muy poco. Se trata del problema siguiente: ¿cómo las ratites, aves sin capacidad de vuelo, se hallan tan ampliamente distribuidas? Las ratites se consideraban antaño aves primitivas, tanto por no poder volar como por hallarse separadas unas de otras por océanos que no podían cruzar. Pareció necesario entonces establecer hipótesis que explicaran la evolución independiente de cada miembro del grupo. El biólogo T. H. Huxley era casi el único que se oponía a esta hipótesis de origen polifilético; basaba su oposición en estudios del paladar óseo de las aves.

Los estudios actuales realizados por Cracraft y por Walter J. Bock, de la Universidad de Columbia (y que se refieren, respectivamente, al esqueleto postcranial y al cráneo de estas aves), indican que las ratites pertenecen a las aves avanzadas y no a las primitivas. Entre sus características distintivas principales se cuenta un paladar de morfología similar al del tinamú, ave parecida a una perdiz y que es nativa de Sudamérica. Puede suponerse, por tanto, que tinamúes y ratites comparten un antepasado común, y que la separación de las ratites del árbol filogenético de las aves voladoras no precedió al surgimiento de los tinamúes.

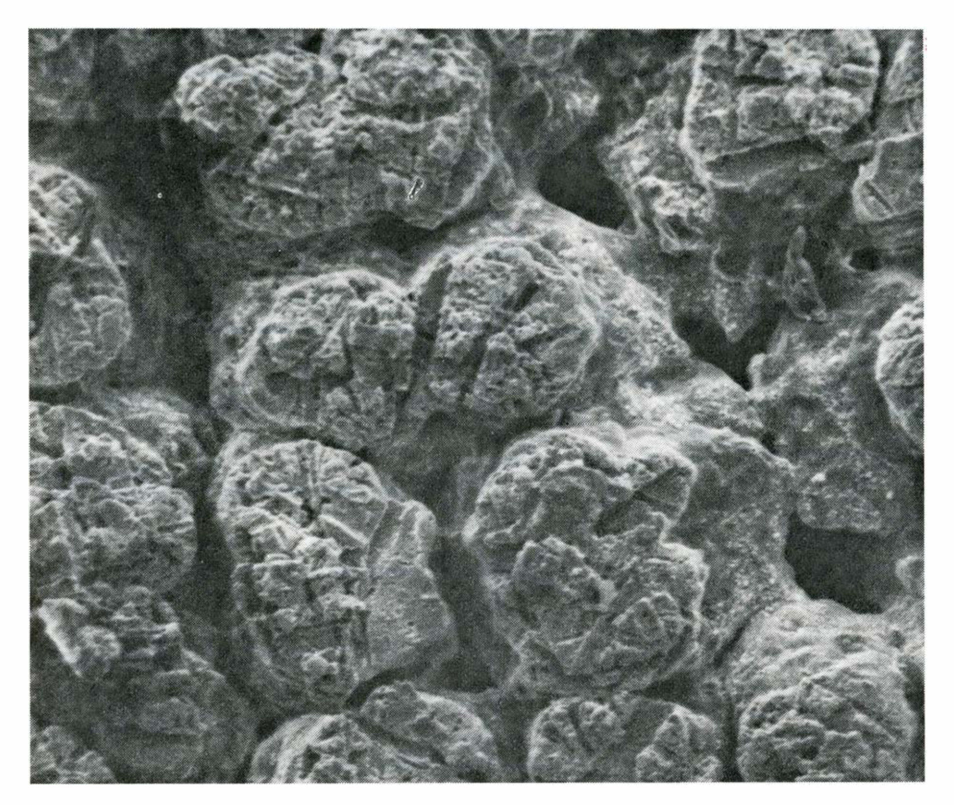
Algunas propiedades de las proteínas de la clara del huevo de las ratites, así como sus relaciones inmunológicas, han sido analizadas por David Osuga, Robert E. Feeney, Ellen Prager y Allan C. Wilson, de la Universidad de California en Davis y de la Universidad de California en Berkeley. Sus hallazgos permiten construir un árbol filogenético de las ratites muy parecido al que había propuesto Cracraft. La única diferencia significativa entre los dos está ligada al hecho de que Cracraft relaciona estrechamente los avestruces con los ñandúes en base a características anatómicas. Otros han señalado que las dos aves comparten piojos del mismo género y que, mientras las ratites del Pacífico tienen únicamente alas vestigiales, el avestruz y el ñandú poseen todavía alas suficientemente grandes para ser utilizadas en las exhibiciones nupciales. Los estudios inmunológicos, sin embargo, sugieren que los ñandúes se encuentran de hecho más cerca del kiwi que del avestruz. Estos hallazgos, sin tener en cuenta que todavía quedan detalles menores por establecer, parecen haber echado a pique la hipótesis del origen polifilético, si se supone, naturalmente, que la curiosa distribución geográfica de las ratites puede explicarse de otro modo.

Cracraft ha señalado que una explicación simple ya la proporciona de hecho la deriva continental. El modelo se corresponde claramente con la disgregación del gran continente austral de Gondwana, que tuvo sus inicios a finales del período Jurásico. Hacia el final del Cretácico, hace unos 90 millones de años, Africa estaba completamente separada de Sudamérica, lo que proporciona la posibilidad de un temprano aislamiento entre ñandúes y avestruces que explique la distancia inmunológica entre estas dos aves. Nueva Zelanda empezó a separarse de Antártica occidental hace unos 80 millones de años, lo que condujo al aislamiento de la ratite ancestral, cualquiera que fuese, que dio origen a los moas y a los kiwis.

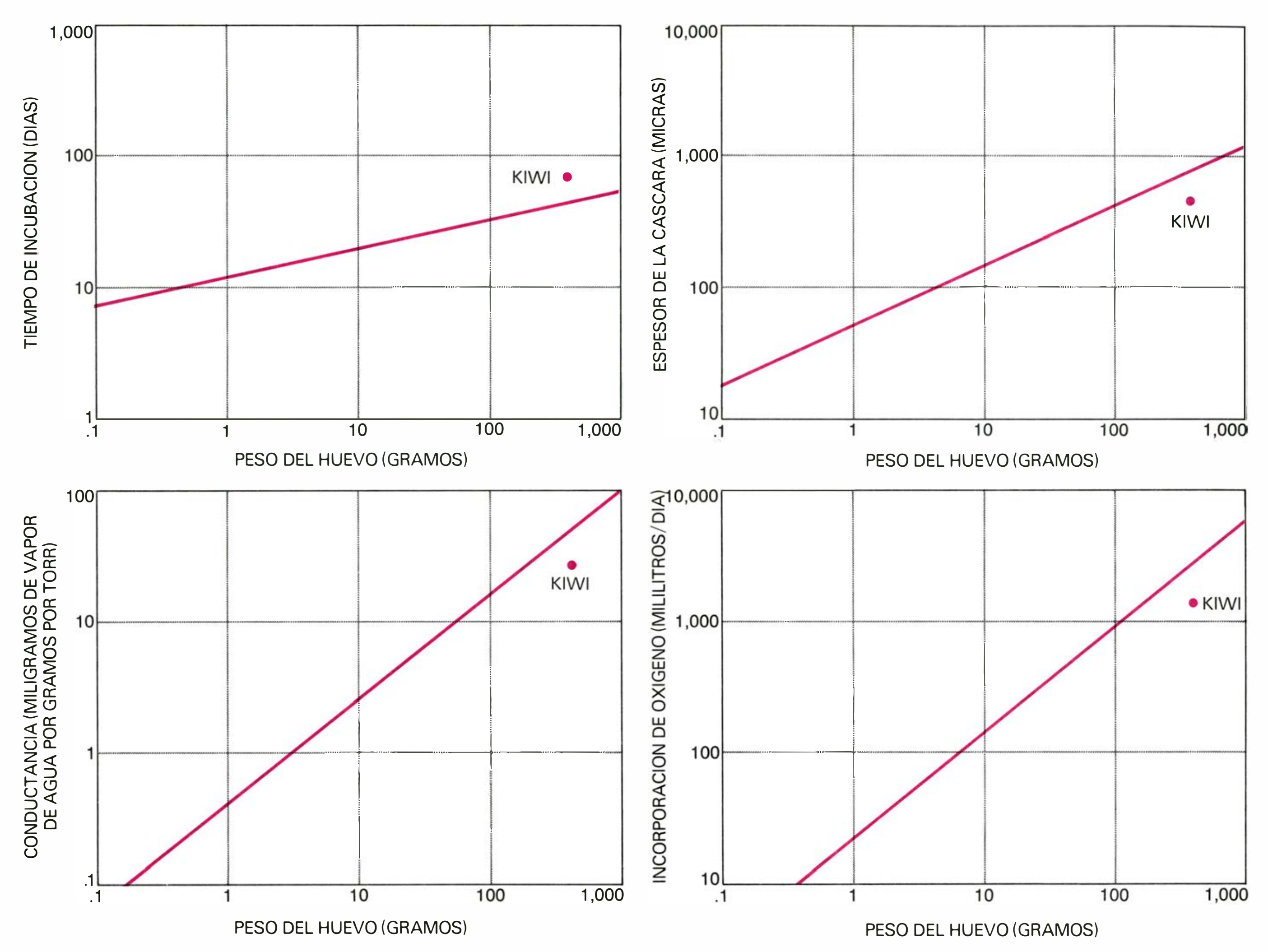
¿Cuándo surgieron los extremos de tamaño entre las ratites de Nueva Zelanda? ¿No será el kiwi quizás un moa encogido? Los únicos restos conocidos de moa son relativamente recientes, de modo que no se sabe nada de la estirpe original de las ratites. Si se acepta la hipótesis monofilética del origen de las ratites, y la verosimilitud de que la capacidad de volar de las aves se perdió en el curso de su aumento de tamaño, la conclusión obligada es que los kiwis son considerablemente más pequeños que sus antepasados.

Todo ello se interpreta en términos de alometría, que Stephen J. Gould, de la Universidad de Harvard, ha definido claramente como "el estudio del tamaño y sus consecuencias". Por ejemplo, si el tamaño de un organismo se aumenta en proporción, no todas sus características anatómicas pueden cambiar a la misma tasa si ha de mantenerse la misma función. Un esqueleto o la cáscara de un huevo han de ser desproporcionadamente más pesados en el organismo mayor para proporcionar un margen de seguridad frente a la tensión de una masa mayor. Por otro lado, si el peso de un organismo se duplica, su tasa metabólica no se multiplica por dos, sino que sólo aumenta en un 68 por ciento, y frecuencias como el ritmo cardíaco o la tasa respiratoria disminuyen de hecho en un 20 por ciento.

El análisis alométrico es simple y gratificador. Ello se debe a que, dentro de una clase determinada de organismos, no sólo las dimensiones estructurales, sino también las tasas y frecuencias de los procesos orgánicos son funciones exponenciales del peso del cuerpo. Cuando las funciones se representan a escala logarítmica, mues-



INTERIOR DE LA CASCARA DEL HUEVO de un kiwi, en esta micrografía electrónica de barrido, que muestra la estructura a un aumento de 325 diámetros. Los cristales de calcita que son secretados por la glándula de la cáscara forman una serie de depósitos aproximadamente circulares, los conos mamilares. La secreción continuada hace que los conos crezcan; los espacios entre los conos son los poros respiratorios de la cáscara del huevo. La micrografía fue realizada por Cynthia Carey, de la Universidad de Colorado.



LAS RELACIONES entre los huevos de las aves son evidentes en las líneas rectas que aparecen en estos gráficos a escala logarítmica que indican aumento en peso del huevo (escala horizontal) frente a cuatro factores determinados por Hermann Rahn y sus colaboradores en la Universidad estatal de Nueva York, en Buffalo. Estos factores son: tiempo de incubación aumentado *(izquierda, arriba)*, grosor de la cáscara aumentado *(derecha, arriba)*, tasa de evaporación desde el huevo aumentada *(izquierda, abajo)* e incorporación de oxígeno aumentada *(derecha, abajo)*. En todos los casos el huevo del kiwi no sigue la regla. Su tiempo de incubación es un 60 por ciento más largo que lo normal, su cáscara es cerca de un 40 por ciento más delgada y sus tasas de difusión de vapor de agua y de incorporación de oxígeno se hallan por debajo de lo normal, aunque poros más cortos de una cáscara que sea más delgada debieran acelerar la difusión.

tran, por lo general, direcciones en línea recta.

El primer estudio alométrico de la relación entre el peso de los huevos de ave y el peso de las aves que los ponen fue publicado por el nieto de T. H. Huxley, Julian. Por ejemplo, un colibrí que pese 3,5 gramos pondrá un huevo de 0,5 gramos, y un avestruz que pese 100 kilogramos pondrá un huevo que pese entre 1,4 y 1,7 kilogramos. En otras palabras, aunque el huevo de avestruz pese mucho más que el de colibrí, será muy ligero en proporción al peso del avestruz. Interpolando alométricamente, un kiwi del tamaño de un gallina debería poner un huevo del tamaño del de una gallina o quizá ligeramente mayor, con un peso entre 55 y 100 gramos. En realidad, el huevo del kiwi pardo pesa, por lo general, entre 400 y 435 gramos. ¿Cuánto debería pesar un ave para poner un huevo de 400 gramos? La ecuación alométrica nos responde que 12,7 kilogramos, cerca de seis veces el peso de una hembra de kiwi.

Los principales estudiosos norteamericanos del análisis alométrico de huevos de aves son Hermann Rahn y sus colegas en la Universidad estatal de Nueva York en Buffalo; Amos Ar, Charles V. Paganelli, Robert B. Reeves, Douglas Wangensteen y Donald Wilson. Su trabajo ha producido ecuaciones que predicen una extraordinaria variedad de propiedades relacionadas a partir del peso de un huevo: el espesor de la cáscara, la permeabilidad de la misma a los gases y al vapor de agua, el período de incubación probable, la máxima tasa de incorporación de oxígeno y la pérdida total de agua durante la incubación. Por ejemplo, la difusión de vapor de agua desde el albumen, o clara del huevo, a través de la cáscara porosa puede medirse como una pérdida de peso a lo largo del período de incubación. La tasa de pérdida puede caracterizar la conductancia (permeabilidad) al vapor de agua del mismo modo que la conductividad o resistencia eléctrica se determina a partir del voltaje y de la corriente de un circuito eléctrico. El oxígeno se difunde hacia el interior a través de los mismos poros, de modo que la incorporación máxima de oxígeno es directamente proporcional a la conductancia de vapor de agua. Por tanto, una vez conocida la tasa de pérdida de agua, puede predecirse el metabolismo del embrión hasta el momento de la eclosión.

El desarrollo del huevo de las aves empieza cuando el óvulo, que consta de una célula germinal y una ración de vitelo nutritivo, se forma en el ovario. El albumen, las membranas y la cáscara del huevo asociados se añaden después de la ovulación, mientras el óvulo pasa a través del oviducto. La proporción del vitelo (yema) con respecto a la masa total del huevo está directamente relacionada con la duración del desarrollo embrionario antes de la eclosión. Las aves altrices (nidícolas), es decir, las que surgen del huevo ciegas y sin plumas, proceden de huevos que quizás en su 20 por ciento son vitelo. Las aves precoces (nidífugas), que nacen

ya cubiertas de plumón y son capaces de alimentarse por sí mismas, proceden de huevos que tienen alrededor del 35 por ciento de vitelo. Dado el gran tamaño del huevo del kiwi y su prolongada incubación (71 a 75 días en lugar de los 44 días que serían normales para un huevo de 400 gramos), no es ninguna sorpresa el hecho de que el joven kiwi nazca completamente cubierto de plumas y no reciba ningún alimento de los padres. Lo que resulta más notorio es la extensión de los preliminares que se esconden tras la precocidad del kiwi. Reid ha encontrado que el contenido del huevo del kiwi tiene un 61 por ciento de vitelo, una proporción que es casi el doble de la que se encuentra en el huevo de las aves precoces típicas.

Una vez se ha depositado la cáscara del huevo, no pueden añadirse a éste más alimento ni más agua. De modo que para ese momento el huevo debe contener suficiente vitelo para nutrir el embrión. Debe incluir asimismo agua suficiente, y la integridad estructural de la cáscara del huevo debe ser tal que el agua se conserve de manera adecuada. Rudolph Drent, que se encuentra en la actualidad en la Universidad de Groninga, en Holanda, ha medido la tasa diaria de pérdida de agua en huevos grandes y pequeños a lo largo de la incubación natural. A partir de los resultados de Drent, y teniendo en cuenta la relación entre el tiempo de incubación y el tamaño del huevo, Rahn y Ar han establecido que el porcentaje de pérdida de agua es muy similar en los distintos tipos de huevos: 18 por ciento para los huevos más pequeños y 13 por ciento para los mayores.

Es excitante observar que existen estas normas generales. Correlaciones tan fidedignas como éstas sugieren vivamente que la selección natural opera dentro de limitaciones físicas bastante estrictas, que en las aves imponen un determinado orden en el tamaño del huevo, en el tiempo de incubación y en detalles estructurales tales como la porosidad de la cáscara. Si la evolución del huevo de las aves se ha hecho, por así decirlo, dentro de un margen reducido de tolerancias de ingeniería, ¿qué ocurre con el kiwi? ¿Se comprenden debidamente las reglas que rigen la variación de escala en la evolución? ¿Pueden ser reales esas reglas si algunas aves resultan no ajustarse a ellas? Con todas estas preguntas en mi mente tomé recientemente un año sabático para visitar Nueva Zelanda y estudiar los kiwis.

Una de las preguntas que iba dando vueltas en mi mente durante el vuelo de 14 horas desde Los Angeles a Auckland tenía que ver con el período de incubación del kiwi, extremadamente largo. ¿Podría ser este período de incubación prolongado el resultado de un contacto físico reducido entre un ave relativamente pequeña y un huevo bastante grande? Estaba deseando registrar las temperaturas del huevo.

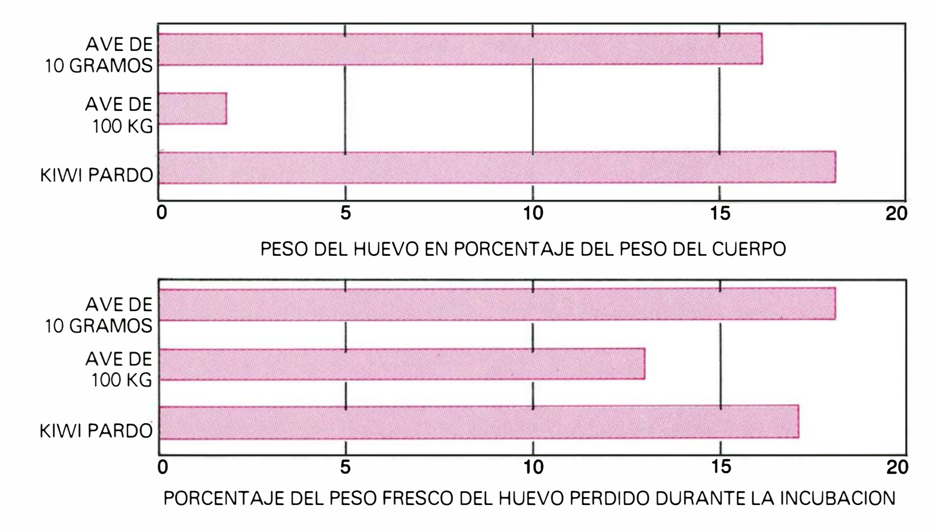
Me había instalado para tomar mi primer refrigerio en Nueva Zelanda, un té en un restaurante situado a orillas del lago Taupo, cuando me llamó la atención un reportaje que estaba transmitiendo la televisión. Barry Rowe, de la Otorohanga Zoological Society, estaba siendo entrevistado sobre su trabajo con los kiwis. Observé incrédulo cómo Rowe abría la puerta de una madriguera artificial de kiwi y extraía un huevo, que abrió en dos mitades para descubrir un pequeño transmisor que emitía un registro continuo de la temperatura del huevo. Después de recobrarme de la sorpresa que representó para mí enterarme de que otro investigador me había ganado por la mano, telefoneé a Rowe. Me invitó a compartir los medios y la hospitalidad de la sociedad zoológica. Me uní rápidamente a Rowe en su trabajo y pronto me hallé atareado midiendo la temperatura de los huevos, que habían sido reforzados y llenados de agua, mientras eran incubados.

Rowe registró los cambios de peso en una hembra de kiwi durante el intervalo entre puestas sucesivas. La acumulación de material que habrá de depositarse en el vitelo rico en energía ocupa los primeros siete días y medio de los 24 que dura el ciclo de puesta. En términos de inversión de energía, durante esta semana la hembra debe aumentar su tasa metabólica basal entre 174 y 203 por ciento. La actividad de alimentación del ave en la naturaleza aumenta probablemente de manera espectacular en el mismo período, pero no hay información cuantitativa acerca del comportamiento del kiwi en libertad que confirme esta suposición.

La hembra de kiwi, que ha invertido una gran cantidad de energía en la síntesis del huevo a lo largo de un período de 24 días, deja de alimentarse en los dos últimos días del ciclo. Como resultado de ello, cuando abandona la madriguera después de haber puesto el huevo, ha perdido más peso que el del huevo. La inversión de energía en el futuro pollo ya ha terminado. La hembra es libre de alimentarse y puede haber entrado ya en un segundo ciclo de síntesis del huevo.

Ahora el macho se encarga de la tarea de incubar el huevo durante cerca de dos meses y medio. En cautividad abandona la madriguera para comer una o dos veces por la noche. Algunos observadores han sugerido que en la naturaleza la hembra también lleva alimento en ocasiones al macho durante este período; hay datos de machos que permanecen en la madriguera sin comer durante varios días.

Al nivel del embrión, en el interior del huevo, la temperatura de incubación alcanza una media de 35,4 grados Celsius. Cuando colocamos un huevo con un sensor térmico bajo una gallina doméstica de raza Bantam (un ave de tamaño inferior al del kiwi), la temperatura de incubación subió a 37,7 grados. Esta diferencia refleja la de la temperatura corporal en las dos aves. Donald S. Farner, de la Universidad de Washington, y sus colegas Norman Chivers y Thane Riney, midieron las temperaturas corporales de varias especies de kiwi. Encontraron que la temperatura basal media del kiwi pardo adulto es de 38 grados Celsius. Esta temperatura es dos

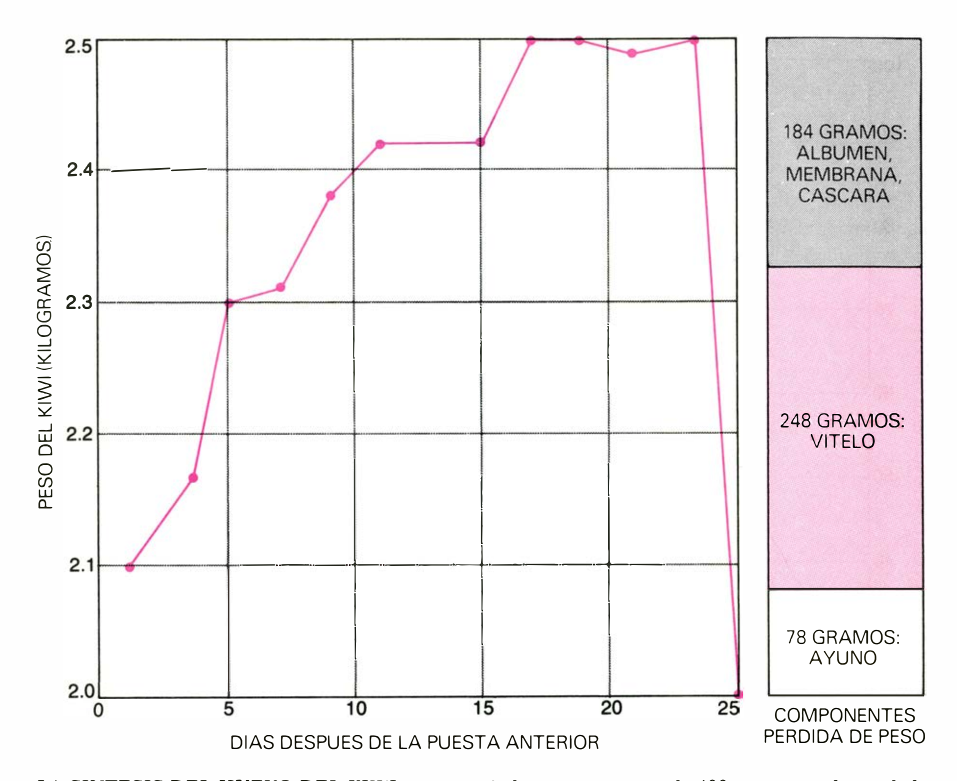


EL PESO DEL HUEVO DE LAS AVES no guarda proporción lineal con el peso de la hembra. La barras del gráfico superior representan el peso del huevo como un porcentaje del peso del progenitor para una pequeña ave teórica que pesa sólo 10 gramos, para un ave grande teórica que pesa 100 kilogramos y para un kiwi pardo, que pesa 2,2 kilogramos. El peso del huevo del ave grande es inferior al 2 por ciento del peso de la madre. Las barras del gráfico inferior representan la pérdida de peso durante la incubación en los mismos tres huevos, expresada como un porcentaje del peso del huevo recién puesto. El huevo pequeño necesitaría incubación durante 13,4 días, el huevo grande durante 62,3 días y el huevo del kiwi durante 71 a 74 días. Las diferencias porcentuales de pérdida de peso no son estadísticamente significativas. Los cálculos se han hecho en base a fórmulas desarrolladas por Rahn y sus colaboradores de la Universidad de Nueva York.

grados inferior a la corporal de otras aves y se parece más a la temperatura de un mamífero. En cualquier caso, si la pequeña gallina Bantam pudo mantener el huevo de kiwi a 37,7 grados, el prolongado período de incubación no puede ser simplemente el resultado de que el huevo sea demasiado grande para que se mantenga caliente.

Conociendo la temperatura de incubación natural del kiwi, Rowe pudo conseguir la primera incubación artificial con éxito de un huevo de kiwi. A lo largo del período de incubación de 71 días, medí la incoporación de oxígeno del huevo a intervalos de dos a tres días. Mientras tanto, al ser difícil la obtención de huevos de kiwi, utilizamos los huevos no fértiles del programa de cría de la sociedad para otras medidas comparadas. El kiwi había probado ser un inconformista no sólo en lo referente al tamaño del huevo, sino también en la temperatura y en la duración de la incubación. ¿Se ajustaba el huevo de kiwi, de algún modo, a las predicciones de Rahn y sus colegas?

Determiné la conductancia al vapor de agua de la cáscara, medí su grosor y separé el vitelo y el albumen. Los elementos separados fueron llevados al Ruakura Agricultural Research Centre, donde C. R. Parr, D. P. Karl y R. Whaanga determinaron el contenido hídrico de ambos y prepararon muestras para la medida del contenido energético del huevo por medio de la bomba calorimétrica. El contenido energético medio de cinco huevos era de 4014 kilojoules (959 kilocalorías). Esta cifra verificaba una estimación anterior de Reid, que había supuesto una semejanza entre el huevo de kiwi y los huevos de las aves domésticas.



LA SINTESIS DEL HUEVO DEL KIWI es acompañada por un aumento de 400 gramos en el peso de la hembra. Durante los primeros siete días del ciclo de 24 se ganan unos 200 gramos. Los requerimientos alimentarios de la hembra deben ser muy elevados durante esta semana. Unos 180 gramos de albumen, membrana y cáscara del huevo se sintetizan al tiempo que el huevo viaja a lo largo del oviducto. Los minerales que necesita la cáscara se encuentran en la dieta o se obtienen a partir de los huesos. Los dos últimos días del ciclo la hembra ayuna. B. Rowe, de la Otorohanga Zoological Society, observó los cambios.

Este resultado significa que, en proporción a su tamaño y a su tasa metabólica, el kiwi va por delante de todas las aves en cantidad de energía invertida en el huevo. Esta enorme reserva de energía mantiene al embrión de kiwi durante su desarrollo prolongado. Además, en el momento en que el pollo de kiwi sale del huevo, sólo se ha consumido poco más de la mitad del vitelo. Lo que queda en el saco vitelino se retrae en el interior del abdomen del pollo y proporciona alimento durante las dos semanas siguientes aproximadamente.

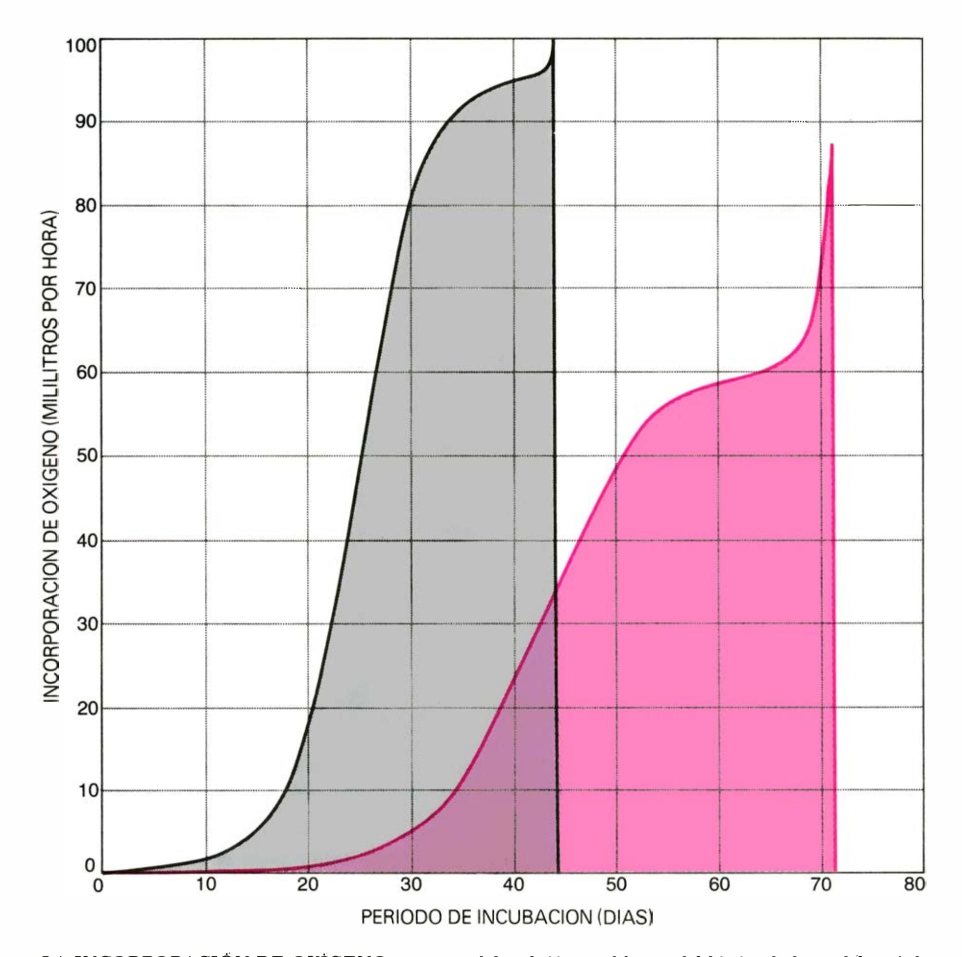
¿Qué ocurre con la pérdida de agua durante el período de incubación de 71 a 74 días? Este período es un 61 por ciento más largo que la predicción alométrica del período requerido para que haga eclosión un huevo normal del mismo peso. Con la evaporación continuada a través de los poros durante un período tan largo, podría esperarse registrar una pérdida de agua que fuera un 61 por ciento mayor si no variaban otras condiciones. Pero encontramos que las demás condiciones variaban.

La cáscara del huevo de kiwi tiene una porosidad muy reducida en comparación con la media de las aves. La reducción proviene de una disminución del número de poros y del tamaño de cada uno de los poros. En consecuencia, la permeabilidad de la cáscara al vapor de agua es de sólo el 60 por ciento de la tasa normal predicha. Esta salida de vapor de agua del huevo, que es más lenta de lo normal, compensa perfectamente el período de incubación, más largo de lo normal. La difusión de oxígeno hacia el interior debe hallarse disminuida de manera similar, naturalmente, puesto que tiene lugar a través de los mismo poros. Efectivamente, esto es lo que indicaban mis medidas de incorporación de oxígeno: la tasa de incorporación era sólo del 63 por ciento de la normal.

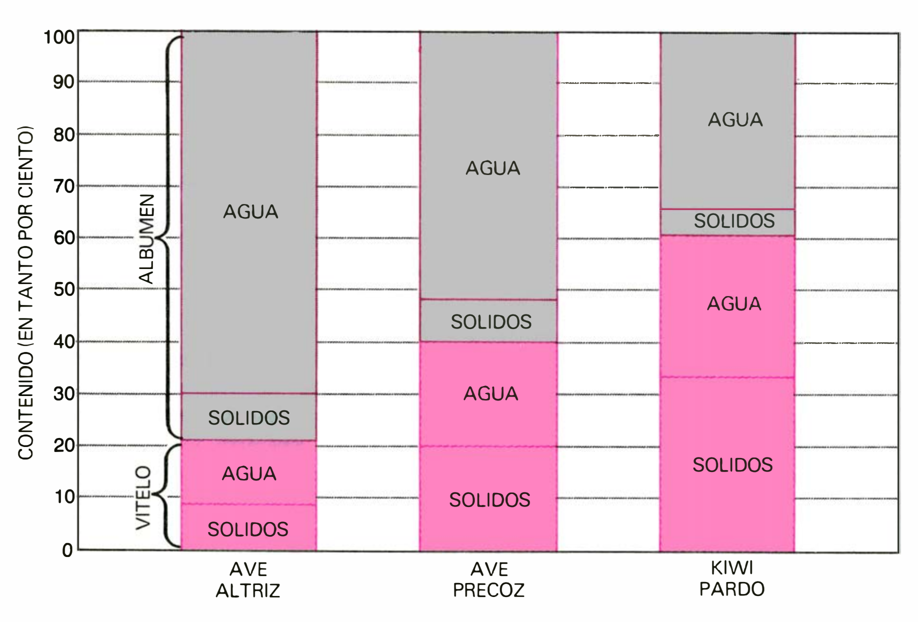
Surge una pregunta: ¿Representa la incorporación de oxígeno más lenta una privación parcial de este gas para el embrión? Se midió el consumo de oxígeno de un pollo de dos semanas de edad que ya había consumido el contenido de su saco vitelino, y también el consumo de oxígeno de dos pollos mayores y de cinco kiwis pardos adultos. Para los pollos, el valor medio de consumo de oxígeno en reposo era del 65 por ciento de la tasa esperada para aves de su peso. Los kiwis adultos promediaron el 61 por ciento de la tasa metabólica prevista. Cuando se probaron de manera similar individuos de las otras dos especies de kiwi, su tasa metabólica resultó ser igualmente baja. De manera que la porosidad reducida de la cáscara del huevo de kiwi limita la incorporación de oxígeno durante la incubación a aproximadamente el mismo nivel que es normal para los kiwis después de la eclosión.

La base para nuestras comparaciones fue la ecuación alométrica para el metabolismo de todas las aves con excepcción de los Paseriformes (el orden que incluye todas las aves canoras). ¿Podría ser que las demás ratites tuvieran asimismo una tasa metabólica inferior a la normal? En colaboración con Terry Dawson, de la Universidad de Nueva Gales del Sur, medimos la tasa metabólica de los emúes. También estos animales resultaron tener una tasa de consumo de oxígeno menor de la que podía esperarse en base a su peso corporal. Quizá debiera reexaminarse el metabolismo basal de todas las ratites. Puede resultar hallarse por debajo de la media de las aves, del mismo modo que en los pájaros está por encima de la media (y en los marsupiales se halla por debajo de la media de los mamíferos placentarios).

Del mismo modo que poner el huevo señala el término de las obligaciones de la hembra, el macho concluye sus obligaciones a la eclosión, a menos que la hembra haya puesto otro huevo. El pollo de kiwi



LA INCORPORACIÓN DE OXÍGENO por parte del embrión en el huevo del kiwi, a lo largo del período de incubación de 71 días *(color)* se compara aquí con la demanda de oxígeno prevista para un huevo normal de 400 gramos *(negro)*. El período de incubación para el huevo normal es de 45 días. Una vez el embrión de kiwi ha pasado el primer mes de incubación, la incorporación de oxígeno aumenta mucho, hasta que pasado el segundo mes alcanza un nivel constante de unos 60 mililitros por hora. Las áreas que encierran las dos curvas son iguales; ello indica que el coste metabólico total del desarrollo, lento o rápido, es el mismo.



SE COMPARA AQUI EL CONTENIDO de tres huevos. Para el huevo y el albumen se distingue el contenido sólido del contenido en agua. El estado de desarrollo de un pollo al salir del huevo está correlacionado con el porcentaje de vitelo en el huevo. El pollo de las aves altrices nace desnudo, ciego y desvalido, de un huevo que tiene hasta un 20 por ciento de vitelo. El pollo de las aves precoces, incubado en un huevo que tiene un 40 por ciento de vitelo, puede correr tras su madre una hora después de haber salido del huevo. El pollo del kiwi pardo no recibe ningún cuidado paterno. En su larga incubación es mantenido por un vitelo que representa el 61 por ciento del contenido del huevo; los sólidos del vitelo son aproximadamente un tercio proteínas y dos tercios grasas. Los sólidos del albumen son, sobre todo, proteínas. Los datos aducidos sobre el contenido del huevo de kiwi fueron aportados por Brian Reid, del New Zealand Wildlife Service.

se alimenta ahora de su saco vitelino hasta que es capaz de alimentarse por sí mismo. Este sistema de ayuda vital funciona muy bien. Por ejemplo, un pollo de kiwi que pesaba 325 gramos inmediatamente después de salir del huevo, bajó hasta 225 gramos antes de empezar a ganar peso otra vez como resultado de alimentarse por sí mismo. La diferencia de peso de 100 gramos correspondía al almuerzo que mamá había preparado casi tres meses antes.

La relación peso del huevo/peso del adulto en el kiwi, ¿representa una disminución evolutiva del tamaño del cuerpo o bien un aumento evolutivo del tamaño del huevo? Una manera de encontrar la respuesta es la siguiente: supongamos que los antepasados del kiwi eran aves voladoras que llegaron a Nueva Zelanda independientemente de los moas y de otras ratites. Entonces, el gran huevo de los kiwis actuales, en lugar de ser el legado de algún antepasado emparentado con los moas, se habría desarrollado por selección natural después de la llegada de estos hipotéticos prekiwis voladores.

¿Cuáles son las ventajas selectivas de un huevo grande? Favorecería un desarrollo embrionario más extenso que el típico de las aves precoces, lo que puede servir para prever demandas de energía según la abundancia natural del alimento. Sin embargo, la cantidad total de alimento en Nueva Zelanda presenta una reducida fluctuación anual, por lo que no existe una necesidad aparente de almacenar gran cantidad de nutrientes en el huevo.

Varias especies de aves alcanzaron volando Nueva Zelanda; desde entonces han perdido o casi su capacidad de vuelo. Entre ellas se cuentan los wekas, dos especies de rascones que ahora son incapaces de volar (*Gallirallus australis* y *G. hectori*); el tacaé, un ave acuática no voladora del grupo de las pollas de agua *(Notornis hochstetteri)*, el pato no volador de la isla Auckland *(Anas aucklandica)* y el cacapú, *(Strigops habroptilus)*, un loro que ya no puede volar aunque todavía planea. Cada ave de éstas, aunque ha evolucionado hacia un estado no volador, ha mantenido las mismas proporciones de tamaño del huevo a tamaño del cuerpo que sus parientes voladores. Esto sugiere que no hay ningún imperativo ambiental en Nueva Zelanda que iguale la incapacidad de volar con el tamaño aumentado del huevo. Además, los huevos de moa, aunque grandes, no lo eran de manera desproporcionada.

Para que el tamaño del huevo aumente debe existir una ventaja selectiva que supere cualesquiera fuerzas que preserven la norma alométrica general. No obstante, si alguna vez existió alguna presión selectiva

para un tamaño menor del cuerpo, por ejemplo en la evolución de un kiwi a partir de un moa ancestral, la retención de un huevo desproporcionadamente grande podría representar meramente la ausencia de una fuerte presión selectiva para la economía en el tamaño del huevo, quizás inducida por las ventajas de que surgiera un pollo más desarrollado a la eclosión.

Las piezas de la historia del kiwi pueden reunirse de más de una manera. Prefiero considerar esta curiosa ave como un ejemplo clásico de evolución convergente. A este respecto, un organismo aviar ha adquirido un notable conjunto de características que, por lo general, asociamos no con las aves, sino con los mamíferos. El archipiélago neozelandés, templado y boscoso, proporciona buenos hábitats para los mamíferos exóticos introducidos. Cuando no había mamíferos para reclamar los nichos que ofrecía este ambiente hospitalario, las aves fueron libres de hacerlo.

El kiwi debe todavía poner huevos; después de todo, es un ave. Con todo, es similar a los mamíferos en varios aspectos. Por ejemplo, Kinsky ha señalado que los kiwis son únicos entre las aves porque retienen ambos ovarios completamente funcionales, de modo que la hembra alterna entre un ovario y otro en las ovulaciones sucesivas, como hacen los mamíferos. También como ocurre en los mamíferos, el largo desarrollo del embrión de kiwi tiene lugar a una temperatura por debajo de lo que es la norma en las aves. El período de incubación de 70-74 días del kiwi está mucho más cerca de los 80 días de embarazo de un mamífero del mismo peso que del período de 44 días que debería bastar para que eclosionara un huevo del tamaño adecuado para un kiwi.

Cuando a esta lista se añaden el hábitat de madriguera del kiwi, sus plumas corporales parecidas a pelaje y su búsqueda nocturna de alimento, que depende en gran medida del sentido del olfato, las pruebas en favor de la convergencia parecen abrumadoras. Yo añadiría a esta lista, y sólo medio en broma, el comportamiento agresivo del kiwi. En el curso de mi investigación en la Otorohanga Zoological Society tenía que entrar con frecuencia en un gran corral que era el territorio de un kiwi macho en período de reproducción. Cuando me introducía en su dominio por la noche, corría hacia mí gruñendo como un gato peleón, agarraba mi calcetín con su pico y clavaba repetidamente sus garras en mis tobillos hasta que me marchaba. Por este comportamiento y por todas las demás razones que he citado, otorgo a esta notable ave la condición de mamífero honorario.

Póker computerizado

Matemáticos, economistas y psicólogos se han interesado por este conocido juego, que proporciona modelos de adopción de decisiones en la vida real. Se está usando ya como vehículo de determinadas investigaciones en ciencias de computación

Nicholas V. Findler

Durante una reciente sesión de póker, uno de los jugadores se dedicó a farolear, exagerando frente a sus oponentes el valor de sus cartas. De esta forma consiguió ganar varias bazas, a pesar de que en casi todas las ocasiones alguno de sus adversarios tenía juego superior al suyo. Tras cierto número de manos, el segundo jugador adaptó su estilo de juego, y acomodándolo al del contrario, comenzó a ganar. Este segundo jugador, cuyo juego fue mejorando en el desarrollo de la partida, era un programa de computador. En la actualidad se están programando máquinas computadoras para que sean capaces de jugar al póker y, al igual que los humanos, aprender por experiencia.

Los juegos ofrecen interés para economistas, psicólogos, políticos, estrategas militares y otra diversidad de gentes, pues constituyen microcosmos donde abundan situaciones reales de adopción de decisiones. Ha surgido una rama de las matemáticas, llamada teoría de juegos, con el fin de determinar las estrategias óptimas a seguir en los juegos o en situaciones abstractas análogas a ellos; situaciones que implican competencia, intereses antagónicos y adoptar decisiones con información incompleta. El póker es, a este respecto, de particular interés, porque en él se dan con suficiente realismo situaciones que exigen adoptar decisiones que conllevan riesgos e incertidumbres.

Entre otros aspectos, el juego de póker presenta el problema de administrar fondos a largo plazo. (Una sesión de póker no dura, por lo común, sino algunas horas, pero en este tiempo es perfectamente posible realizar 100 o más partidas, en cada una de las cuales se presentan al menos una docena de situaciones comprometidas.) Por consiguiente, el juego puede aportar información sobre muchos problemas distintos (como en planificación económica, campañas electorales y confrontaciones bélicas) en los que cierta magnitud cuantificable (que en estos ejemplos podrían ser recursos financieros, votos u objetivos militares) ha de hacerse lo mayor posible al cabo de un tiempo considerablemente largo.

El póker comparte con estas otras áreas de la actividad humana una notable característica. En el póker es preciso tomar decisiones fundándose en informaciones incompletas. Comparemos el póker con el ajedrez. Este último es un juego de información perfecta, es decir, cada jugador conoce todos los elementos de las situaciones pasadas y presentes; el problema de los jugadores es realizar en un tiempo reducido los complejos cálculos requeridos para determinar un movimiento satisfactorio. En el póker ningún jugador tiene información completa de las situaciones pasadas y actuales. Parte de la información de que cada uno dispone es segura (por ejemplo, las propias cartas, o el historial de apuestas del decurso del juego), pero otra parte es arriesgada (por ejemplo, la probabilidad de recibir ciertas cartas tras el descarte) y aún otra es insegura (por ejemplo, la conducta futura de sus adversarios). Los jugadores han de formarse idea de las cartas de sus adversarios, labor dificultada todavía más por estar permitidos los faroles, que son parte esencial del juego. Los jugadores han de sintetizar diversidad de informaciones de varios tipos, y recurrir a un modelo psicológico de sus adversarios, que es necesario corregir continuamente, para estimar la propia situación en el juego.

Los procesos de decisión de este tipo no son nada fáciles de reproducir mediante programas de computador, pues descansan fuertemente en el juicio humano. Sin embargo, justamente por esta dificultad es la programación del póker un problema interesante para los estudiosos de las ciencias de computación. De resolverse los problemas intrínsecos de los programas de póker podrían aclararse bastante otros muchos problemas que dependen de la perspicacia e intuición humanas. En particular, si se consiguiera comprender más profundamente el juego de póker, tal información podría aplicarse en otros muchos procesos de decisión

Mis discípulos y yo estamos actualmente trabajando, en la Universidad estatal de Nueva York en Buffalo, en un proyecto encaminado a explorar los procesos cognoscitivos implícitos en el póker. Se están desarrollando experimentos con seres humanos, a fin de definir con precisión suficiente los procesos de toma de decisiones aplicados en el transcurso del juego; al mismo tiempo, se realiza un gran esfuerzo para poner a punto técnicas de programación satisfactorias. En este artículo describiré el tema principal de nuestra investigación: la ppreparación de estrategias de máquina, o programas de cmputador que puedan jugar al póker de acuerdo con diversos principios. Con ese objeto comenzaré esbozando la forma de póker utilizada en nuestra investigación: el póker con descarte.

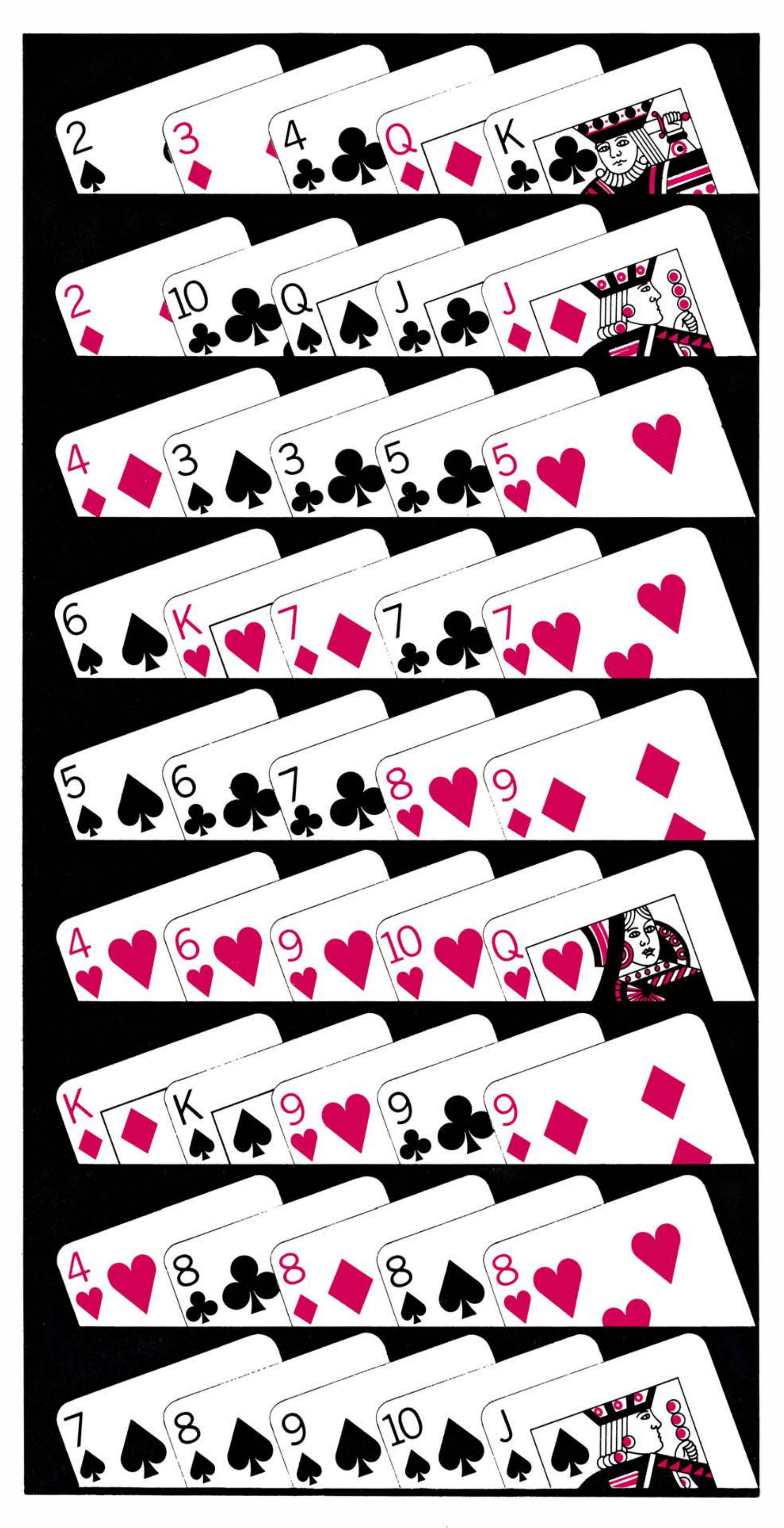
En el póker con descarte, cada partida se inicia pagando cada jugador una puesta de entrada (llamada "ante"), a cambio de la cual recibe una mano de cinco cartas, que oculta a sus adversarios. Una vez que los jugadores examinan sus cartas, el primer jugador situado a la izquierda del que ha repartido tiene la opción de abrir el primer período de apuestas del juego. (En la mayoría de las versiones del póker, el primer apostante inicial ha de tener, cuando menos, una pareja de sotas ["jacks"]; de no ser así, la opción de apertura va pasando correlativamente alrededor de la mesa en el sentido de las agujas del reloj.) Una vez realizada la apuesta de apertura del juego, el turno de apuestas va corriendo alrededor de la mesa en sentido horario. Llegada su vez, cada jugador ha de resolverse por una de estas tres alternativas: pasar (abandonando el juego y perdiendo en consecuencia la puesta inicial), entrar (añadiendo dinero suficiente para que su aportación total al monto iguale a la del último jugador que tomó parte activa) o subir (haciendo una nueva apuesta, con aportación de más dinero del que sería necesario para entrar).

La fase de apuestas concluye cuando cada jugador activo ha puesto en total la misma cantidad de dinero en el centro de la mesa. (El número de rondas de apuestas de cada período está limitado por una norma estipulada al principio de la sesión de póker.) La siguiente fase del juego es el descarte, en el cual los jugadores aún activos pueden desprenderse de una o más cartas de su mano, recibiendo a cambio nuevas cartas del mazo de naipes no repartido. A continuación se abre un segundo período de apuestas, que puede concluir de una de estas dos maneras: Si pasan todos los jugadores excepto uno, este último gana el juego. O si todavía quedan varios jugadores en activo al concluir la segunda fase de apuestas, entonces se vuelven las cartas boca arriba: cada jugador en activo muestra su mano, siendo ganadora la de mayor fuerza.

Las decisiones que ha de tomar un jugador de póker a lo largo del juego se basan en cierto número de criterios, de los cuales el más importante es el valor de su mano. En muchos casos, sin embargo, los jugadores actúan de formas que no reflejan fielmente, por así decirlo, su posición en el juego o la fuerza de la mano que poseen. Tal comportamiento se llama "faroleo" ("bluffing") y puede utilizarse con diferentes propósitos. Un jugador puede vencer en el juego exagerando una mano



LOS JUEGOS DE POKER en que participan programas de ordenador pueden visualizarse en la pantalla del tubo de rayos catódicos de un terminal de computador. La ilustración muestra el formato presentado en la pantalla, que permite que un jugador humano interactúe (en este caso, compita en un juego de póker de descarte) con un jugador programado (programa). Un programa de juego es un programa de ordenador, diseñado para jugar al póker de acuerdo con una determinada estrategia. Los siete pequeños cuadros azules que rodean el círculo obscuro de la pantalla representan programas de juego; el cuadrado rosa situado bajo la mano de naipes simboliza un jugador humano, quien se sentaría frente al terminal. La mano de naipes pertenece al jugador humano. También puede verse en la pantalla la situación actual del juego. El juego se encuentra en la fase previa al descarte. El monto (o total de apuestas) asciende a 18 tantos. El jugador 4 repartió las cartas, el jugador 6 abrió las apuestas, con 10 tantos, y el jugador 9 pasó. (Los otros ocho tantos del monto son "antes" –cuotas de entrada–, uno por cada jugador). El próximo será el jugador humano, que posee una pareja de reyes. El jugador puede "pasar" *(fold),* "ir" *(call)* o "subir" *(raise),* apuntando con un lápiz luminoso al cuadro rosa oportuno, situado en la parte inferior derecha. Puede indicar la cuantía de la subida marcándola en la escala de la parte baja de la figura. Tras actuar el jugador humano, el turno de juego va pasando en sentido de las agujas del reloj alrededor del círculo obscuro. Los jugadores de la parte media superior, a la izquierda de la ilustración, Exploit, Asprate y Caller 2, deben su nombre peculiar a las estrategias a que obedecen. Se utilizan formatos de pantalla semejantes para supervisar situaciones de juego donde sólo compiten jugadores de máquina.



RANGO DE VALORES de las manos de póker. Desde la más baja *(arriba)* a la más alta *(abajo)*, el rango de valores es el siguiente: carta más alta (la carta de máximo valor de la mano), una pareja, doble pareja, trío, escalera (cinco cartas de valores correlativos, pertenecientes a dos o más palos), color o flux (cinco cartas cualesquiera de un mismo palo), full (una pareja más un trío), póker (cuatro cartas del mismo valor) y escalera de color (cinco cartas que sean de valores correlativos y que pertenezcan a un mismo palo).

floja, y convenciendo a todos sus oponentes de que deben pasar. O bien, disimulando la auténtica fuerza de sus cartas, evitar que los demás pasen, con objeto de aumentar la ganancia esperada.

En el póker, al igual que en muchos tipos de transacciones comerciales, la decisión de cuánto se debe farolear en una situación dada y con un historial dado es materia de extraordinaria importancia. Si se farolea demasiado o muy poco (es decir, eligiendo torpemente la intensidad y la frecuencia de los faroles), la consecuencia puede ser el desastre financiero. Tenemos al jugador de póker que podríamos denominar "totalmente racional", que se basa en una estrategia matemáticamente equilibrada. Este jugador se abstiene por completo de farolear; para decidir cuánto aportar en un momento dado del juego lo que hace es igualar los valores esperados de sus ganancias y sus pérdidas; es decir, forma una ecuación situando en un miembro el valor esperado de las ganancias y, en el otro, el valor esperado de las pérdidas, y a partir de ella calcula el montante de su apuesta. (El valor esperado, o esperanza matemática, de un suceso es el producto de la probabilidad de que tal suceso ocurra por el valor asociado a él.) El problema estriba en que esta estrategia "racional" solamente podría tener éxito en el caso de que los jugadores contrarios fueran tan inexpertos que no la descubrieran, o en el caso de que todos los demás jugadores fuesen igual de "racionales". En los demás casos, los contrarios del jugador racional no tendrían sino que aumentar sus apuestas, independientemente de la fuerza de sus cartas; llegará un momento en que el jugador racional deberá pasar, perdiendo así la totalidad de su contribución a la partida.

Como muy bien muestra el ejemplo de la estrategia matemáticamente justa, una estrategia de póker satisfactoria ha de contener una juiciosa proporción de faroles y aportar medios para detectarlos en los contrarios. Los faroles se utilizan no sólo para ganar en una partida concreta, sino también para obtener información y para evitar que los oponentes la obtengan. Por ejemplo, un jugador puede ir de farol toda la partida –es decir permanecer en el juego sabiendo que probablemente perderá– con el fin de ver la mano de un adversario. (En el póker de descarte el jugador vencedor no está obligado a mostrar su mano en caso de que todos sus oponentes hayan pasado.) El jugador que va de farol está en este caso comprando información, pagando con el fin de ver la relación entre la fuerza de la mano de sus oponentes y su comportamiento a la hora de apostar. Dicho de otra forma, en contrapartida al dinero que pierde, el jugador farolero recibe

información relativa al estilo de juego de uno o más de sus contrarios. Con faroles de diversos tipos, un jugador puede "vender", o sea, pasar informaciones que despisten a los contrarios respecto del estilo propio. (Este tipo de farol se podría llamar también falsa publicidad.) La compra y venta de información, tan característica del póker, se encuentra también en otras muchas actividades donde se precisa adoptar decisiones.

Conseguir un programa de computador capaz de jugar bien al póker exige primero comprender los procesos cognoscitivos que los seres humanos aplican a este juego. (La teoría matemática de juegos solamente permite analizar versiones simplificadas.) En general, cuando una persona ha de tomar una decisión construye para sí (consciente o insconscientemente) un modelo del problema de decisión. En él se describen los objetos y sucesos ligados al problema, presentando sus significados e interacciones en forma abstracta y simplificada. El proceso de construcción del modelo puede contemplarse como un proceso de progresivo aprendizaje, porque cuando se recibe nueva información el modelo se amplía o se modifica en consecuencia. O si la nueva información está en fuerte desacuerdo con el modelo, puede ser necesario crear uno nuevo. Esta clase de modelos permiten que la mente humana, que es un procesador de capacidad limitada, interprete correctamente el entorno y tome en relación con él decisiones válidas.

La construcción del modelo asociado al juego puede dividirse a grandes rasgos en tres etapas. En la primera, el jugador practica y se familiariza con las reglas. Al hacerse un esquema mental del juego y del problema que presenta, el jugador adquiere también algunas estrategias que, aunque rudimentarias, son mejores que el juego al azar. El proceso de aprendizaje puesto en acción durante esta etapa debería producir (1) un conjunto de rutinas de evaluación, esto es, procedimientos para discernir la fuerza de la mano de los contrarios, su estilo de juego, etcétera, (2) la adquisición de información, y en particular, reglas de juego, rutinas a seguir en situaciones frecuentes y la probabilidad de ciertos sucesos ligados al juego, (3) estructuras adecuadas para almacenar información relativa a la partida que se está desarrollando y (4) rutinas de adopción de decisiones, con el fin de combinar entre sí las informaciones de los tipos anteriores y llegar a establecer un plan de acción.

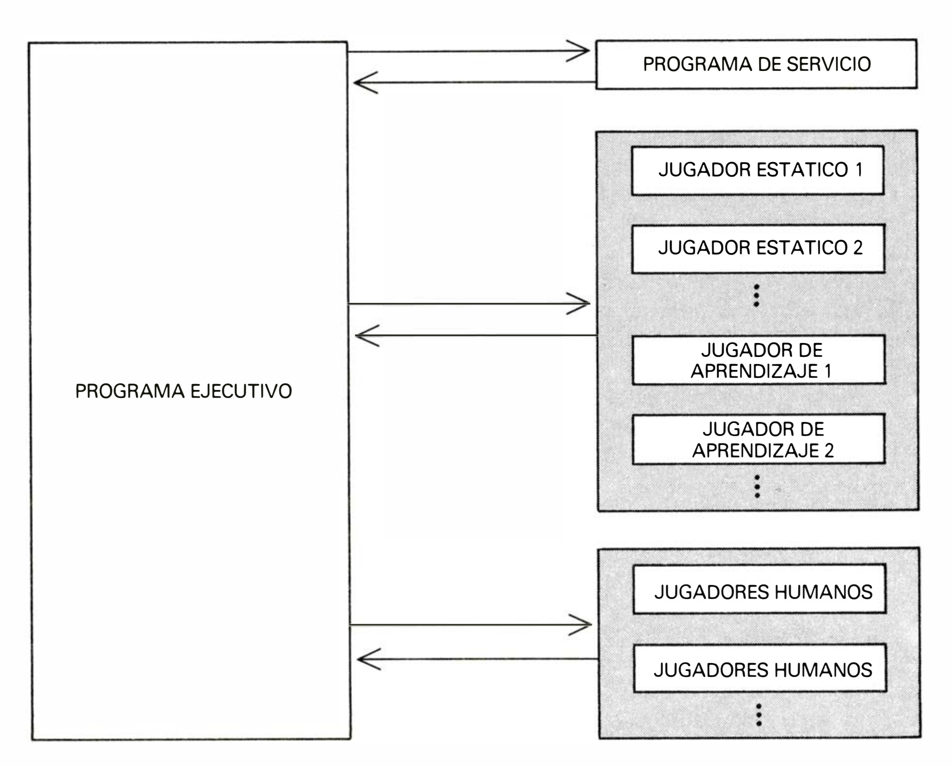
En la segunda etapa de juego, el jugador aprende a generar buenas jugadas, formulando para ello una serie de posibles jugadas que contrasta mediante un sistema de criterios. La revisión del historial del juego y la comparación de los resultados esperados de sus jugadas con los resultados realmente obtenidos permiten que el jugador corrija y mejore sus estrategias.

La anterior descripción del proceso de juego deja sin respuesta muchas cuestiones concernientes a la forma de preparar y aplicar reglas de resolución de problemas en una partida entre jugadores humanos. ¿Cuáles son los criterios de decisión empleados en el póker, y cuál es su importancia relativa? ¿En qué forma puede un jugador evaluar correctamente la situación propia y la de sus oponentes? ¿Qué sucesos o qué configuraciones del juego inducen a los jugadores a cambiar de estrategia? ¿Cuándo está un jugador dispuesto a comprar información y cuánto estará dispuesto a pagar por ella? ¿De qué forma influye en las propias decisiones la imagen que cada jugador se crea de sus contrarios?

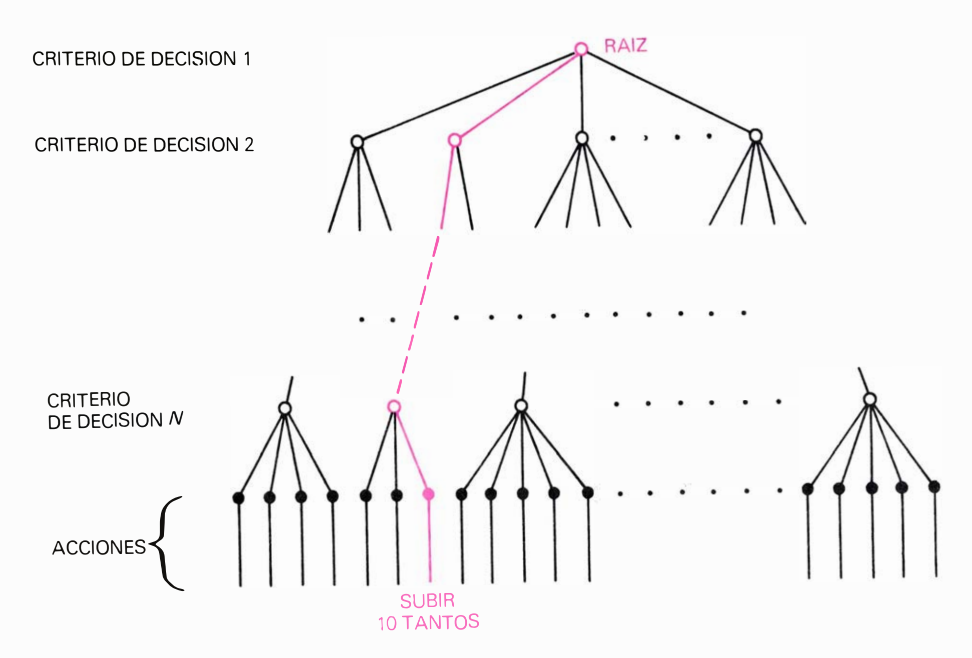
La serie de experimentos que mis discípulos y yo hemos venido realizando fue diseñada con el propósito de solventar estas cuestiones. Nuestros sujetos fueron expertos jugadores de póker, que, al igual que en muchos experimentos de psicología, recibieron una bonificación por participar en el proyecto. Los sujetos juegan al póker arriesgando sus gratificaciones, por lo que los experimentos, aún realizados en condiciones controladas, de laboratorio, implicaban un riesgo financiero real. (Se obtuvo autorización del Departamento de Policía de Buffalo y del comité de ética de la Universidad.) Por consiguiente, los observadores pueden formular generalizaciones válidas sobre los riesgos asumidos en distintas actuaciones.

La descripción detallada del desarrollo de estos experimentos, que prosiguen, rebasa los límites de este artículo. Muchos de los descubrimientos obtenidos han sido ya incorporados a diversas estrategias de máquina. Precisamente me gustaría ahora dedicar mi atención a esta parte del proyecto: la programación de computadores para jugar al póker.

El sistema de programación que hemos concebido se compone de tres partes interactivas. Un programa ejecutivo se encarga del flujo de información y de la transferencia del control entre los diversos componentes. Un gran sistema de programas de servicio realiza tareas de comprobación de errores, recogida y preparación de datos estadísticos, cálculo de valores probables, gestión de las sesiones de juego, etcétera. Finalmente, el sistema se completa con una colección creciente de estrategias de máquina, o sea, programas de juego.



EL SISTEMA DE PROGRAMACION diseñado para el póker computerizado consta de tres partes. El programa ejecutivo controla el sistema y gestiona el flujo de información y la transferencia de control *(flechas)*. Los programas de servicio realizan los diversos trabajos de mantenimiento del sistema, incluida la compilación y cálculo de datos estadísticos, depuración de contradicciones en el programa *(debugging)*, la comprobación de errores de juego y la gerencia de las sesiones. Los programas de juego o jugadores programados son los que determinan la conducta a seguir en las partidas. Hay dos tipos de jugadores programados: jugadores estáticos, cuya estructura y contenido no varían en el desarrollo de la sesión, y jugadores de aprendizaje, cuya estructura y contenido van modificándose conforme adquieren experiencia. Los jugadores humanos pueden interactuar con los jugadores de máquina a través del programa ejecutivo.



UNA ESTRATEGIA DE MAQUINA prescribe la adopción de distintas actuaciones, según los valores de un conjunto de variantes, o criterios de decisión, asociadas a la situación de la partida en cada momento, al historial del juego y a la personalidad de los jugadores. Una de las diversas formas de estructurar el proceso de obtención de decisiones mediante una estrategia de máquina consiste en utilizar un grafo llamado "árbol de decisión". En el árbol de decisión que aquí se muestra cada nivel de nodos está asociado a una variable de estrategia representada por el árbol. Las ramas descendentes que parten de los nodos indican subgamas de los valores que pueden tomar la variable asociada al nodo. Las "hojas" que cuelgan del último nivel de nodos corresponde a acciones del juego. Así pues, cuando se da entrada por la raíz del árbol a una situación de juego, los valores de las variables determinan entonces un cambio hasta la acción oportuna *(color)*.

Estas estrategias de máquina no son colecciones de las estrategias conocidas por el nombre de estrategias puras de póker, es decir, estrategias que prescriben una acción específica para cada situación imaginable. El concepto de "situación" en el desarrollo de cualquier juego, determinada por factores variables que van desde el historial de apuestas y la personalidad de los contrincantes hasta la colocación de ciertos jugadores en la mesa de juego, es una noción tan amplia que la mera enumeración de todas las condiciones de tan sólo una estrategia pura es problema de enorme complejidad. Además, en el póker, lo mismo que en otros muchos juegos, existen muchísimas estrategias puras. (La dificultad de formular estrategias puras, y el enorme número de tales estrategias, explica la imposibilidad práctica de hallar una estrategia óptima para el póker, el bridge, el ajedrez y otros juegos semejantes.)

Las estrategias de máquina que se describirán prescriben jugadas específicas en respuesta a ciertas variables asociadas a una cierta situación de juego, a su historial y a la personalidad de los contrincantes. Cada estrategia está formada por un conjunto ordenado de elementos de decisión, cada uno de los cuales tiene especial relevancia en una etapa concreta del juego. Existen diversas formas de representar estrategias mediante computador. Por ejemplo, ciertas estrategias de máquina se representan en forma de conjunto de reglas de producción. Cada regla consiste en un esquema, o una sucesión de símbolos, que caracteriza a una configuración de elementos que podrían ocurrir en un juego, asociados con una acción a adoptar en la correspondiente situación de juego. Cuando se utiliza una tal representación se genera un patrón que describe la situación actual del juego y de las personalidades de los jugadores, y después tal patrón se compara con los dados en las reglas de producción. Cuando se encuentra una concordancia (aunque sea aproximada) entre el patrón actual y uno de los pertenecientes a las reglas de producción, el programa efectúa la transferencia de control al procedimiento asociado a tal regla particular.

Otro método de representación de procesos de decisión consiste en un árbol de decisión. Un árbol es un grafo que carece de nodos (puntos) aislados, y que carece también de circuitos cerrados. Los nodos del árbol de decisión están organizados en varios niveles, con un único nodo, llamado raíz, en la cima del árbol. Hay ramas que conectan cada nodo con uno perteneciente al nivel superior inmediato y con uno o varios nodos del nivel inmediatamente inferior. Los nodos de un nivel determinado corresponden a una variable del juego asociada a la estrategia representada por el árbol (por ejemplo, el valor de la mano de un jugador o el número de jugadores todavía activos en la partida). Las ramas que descienden de cada nodo corresponden a distintas subgamas de valores que puede tomar la variable asociada al nodo. De los nodos terminales del nivel más bajo del árbol cuelgan "hojas" solas, que representan acciones a realizar en el juego. Por la raíz del árbol de decisión se da entrada a una situación de juego de personalidades de los jugadores, y los valores que en ese momento toman las variables del juego determinan un camino hasta una hoja determinada, es decir, hasta una acción a ejecutar.

También pueden representarse las estrategias de computador mediante una red asociativa: un grafo en el que los nodos no están organizados en niveles jerárquicos. Ahora pueden estar interconectados dos nodos cualesquiera; además, de cada nodo puede emanar un número arbitrario de arcos o líneas de enlace. Los nodos pueden disponerse en distintos "planos" en el caso de que la información que contienen sea clasificada según ciertos criterios como, por ejemplo, las fases del juego a las que estén asociados. En la red asociativa cada nodo está en correspondencia con un cierto elemento del juego (por ejemplo, el número de jugadores todavía activos, o la acción a realizar) y cada arco porta un rótulo que caracteriza la relación entre los nodos que interconecta. Se da entrada a cada situación del juego por un nodo especial, y a partir de allí se va recorriendo el grafo, de nodo en nodo, hasta alcanzar alguno que se corresponda con una acción a ejecutar. Los rótulos de los arcos sirven como indicadores de ruta para atravesar la red, es decir, ayudan a seleccionar el nodo activo que en ese momento se adapta mejor a la situación de juego considerada. Puede demostrarse que los tres métodos de representación de estrategias de computador son equivalentes, aunque, para una estrategia dada, una de ellas puede ser más eficiente que las demás.

Las distintas estrategias de máquina no solamente se diferencian en su estructura, sino también en el método de abordar el problema de tomar una decisión. Algunas de ellas adoptan decisiones fundándose solamente en la inteligencia de la máquina, es decir, obedeciendo reglas lógicas y matemáticas. Otras imitan los métodos humanos de toma de decisiones, apoyándose en recomendaciones de libros de póker y en hallazgos experimentales sobre jugadores humanos. En realidad, la mayoría de las estrategias toman cuando menos algunos elementos de los procesos de decisión realizados por humanos, procesos que se caracterizan por aplicar criterios de decisión poco definidos y por aplicar los llamados razonamientos heurísticos.

Los criterios de decisión mal definidos son los enunciados de modo impreciso o que se limitan a dar una descripción

cualitativa de carácter general. Por ejemplo, un jugador que quisiera "vender" de sí mismo una imagen de jugador timorato podría proyectar unos cuantos abandonos en las fases iniciales de juegos con apuestas fuertes. A pesar de que no es posible atribuir valor exacto al calificativo "iniciales", sí es posible preparar programas de computador que describan cómo dar respuesta óptima a contrincantes que exhiban determinados tipos de conducta.

El razonamiento heurístico, basado en lo que comúnmente se llama perspicacia o intuición, permite que los seres humanos obtengan soluciones de ciertos problemas sin tener que examinar la totalidad de una masa de información (posiblemente enorme) referente a ellos. En un programa heurístico se resuelven los problemas, no mediante la búsqueda exhaustiva de soluciones, sino utilizando ciertas normas prácticas y los diversos atajos y aproximaciones tan característicos de los procesos mentales humanos. Analizando la conducta de jugadores expertos es posible formular reglas heurísticas que luego pasen a formar parte de programas de computador. Los programas heurísticos de este tipo dan soluciones que, si bien no son óptimas, por lo general son francamente buenas y razonablemente eficientes.

Los programas de juego que mis discípulos y yo hemos preparado pueden dividirse en dos categorías, a saber: estrategias estáticas y estrategias de aprendizaje. En las primeras, ni el contenido ni la estructura de la estrategia varían con el tiempo; mientras que en las estrategias de aprendizaje, ya el contenido, ya la estructura, o ambos, van cambiando conforme se acumula mayor experiencia.

No es verdad que una estrategia estática responda forzosamente con acciones iguales a configuraciones idénticas de elementos del juego; se pueden programar faroles, ya sea para realizarlos al azar o en respuesta a determinadas situaciones del juego. La única estrategia estática que no utiliza "faroles" es la del llamado "Jugador Matemáticamente Equilibrado". Obedeciendo al estilo del jugador humano "totalmente racional", que ya ha sido descrito, este programa decide cuándo y qué cantidad apostar equilibrando los valores esperados de ganancias y pérdidas; o sea, dicho de otra forma, sus apuestas se ciñen estrictamente a sus posibilidades de ganar con una mano dada. Evidentemente, esta estrategia no ofrece defensa frente a oponentes que faroleen de un modo sistemático, pero se comporta bastante bien en partidas contra jugadores estáticos no demasiado perspicaces. De hecho, la estrategia del Jugador Matemáticamente Equilibrado resulta conveniente para arrastrar con ella la habilidad de otros jugadores de estrategia estática.

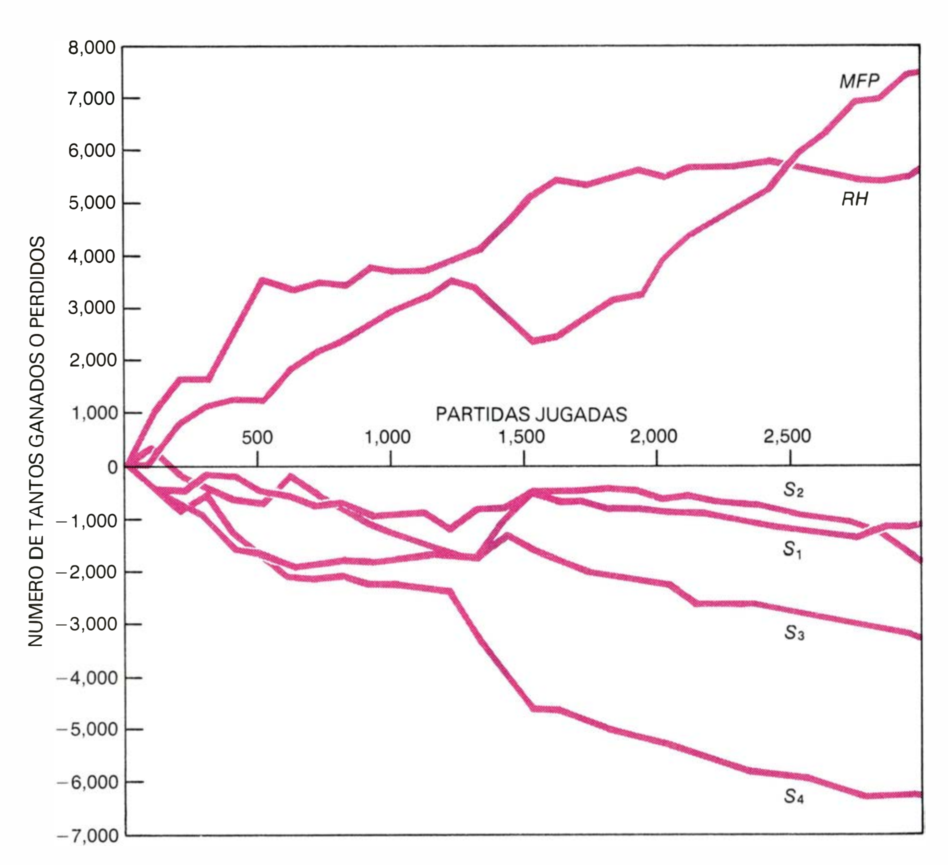
Otra estrategia estática, la del Jugador RH, elabora sus decisiones fundándose en una gama más amplia de informaciones que el Jugador Matemáticamente Equilibrado. El Jugador RH no sólo toma en consideración la fuerza de su mano, sino también los valores de los llamados factores de segundo orden del juego de póker, factores mucho menos importantes que el valor de las cartas, pero que, no obstante, tienen un importante papel a la hora de decidir el resultado del juego. El valor obtenido combinando adecuadamente dichos factores de segundo orden se llama índice RH. (Esta variable fue diseñada por dos de mis discípulos postgraduados, Jean Rachlin y Gary Higgins, de los cuales recibe su nombre.)

El índice RH es una medida del grado de apetencia que un determinado jugador tiene en un momento dado de permanecer en el juego. Para conseguirlo, se pone en el numerador de una fracción la variable que expresa motivación positiva, y en el denominador el producto de las variables de motivación negativa. La motivación positiva es la cantidad de dinero a ganar, que se expresa mediante una variable denotada TOTAL. Todas las motivaciones negativas se refieren a la posibilidad de perder. Entre estos factores está la variable VIVOS, que es igual al número de jugadores todavía en activo. La variable CUENTA es igual al número de veces que se ha subido el monto total de apuestas en el período de apuestas en curso; esta variable puede indicar, entre otras cosas, hasta qué punto están decididos los jugadores en activo a permanecer en el juego. La variable SIGUIENTES es igual al número de jugadores que pueden actuar después de que el Jugador RH tome algún tipo de acción. Finalmente, la variable PLUS es igual al costo de permanecer en el juego, es decir, la cantidad de dinero que el jugador debe aportar para "ir". Así pues, el índice RH es igual a:

$$\frac{\text{TOTAL}}{\text{VIVOS} * (\text{CUENTA} + 1) * \text{SIGUIENTES} * \text{PLUS}}$$

(Se usa en el denominador CUENTA + 1 para evitar que pueda ser igual a cero.)

Rachlin e Higgins efectuaron una serie de experimentos mediante computador para determinar la distribución característica de valores del índice RH. Con ayuda de la máquina generaron aleatoriamente

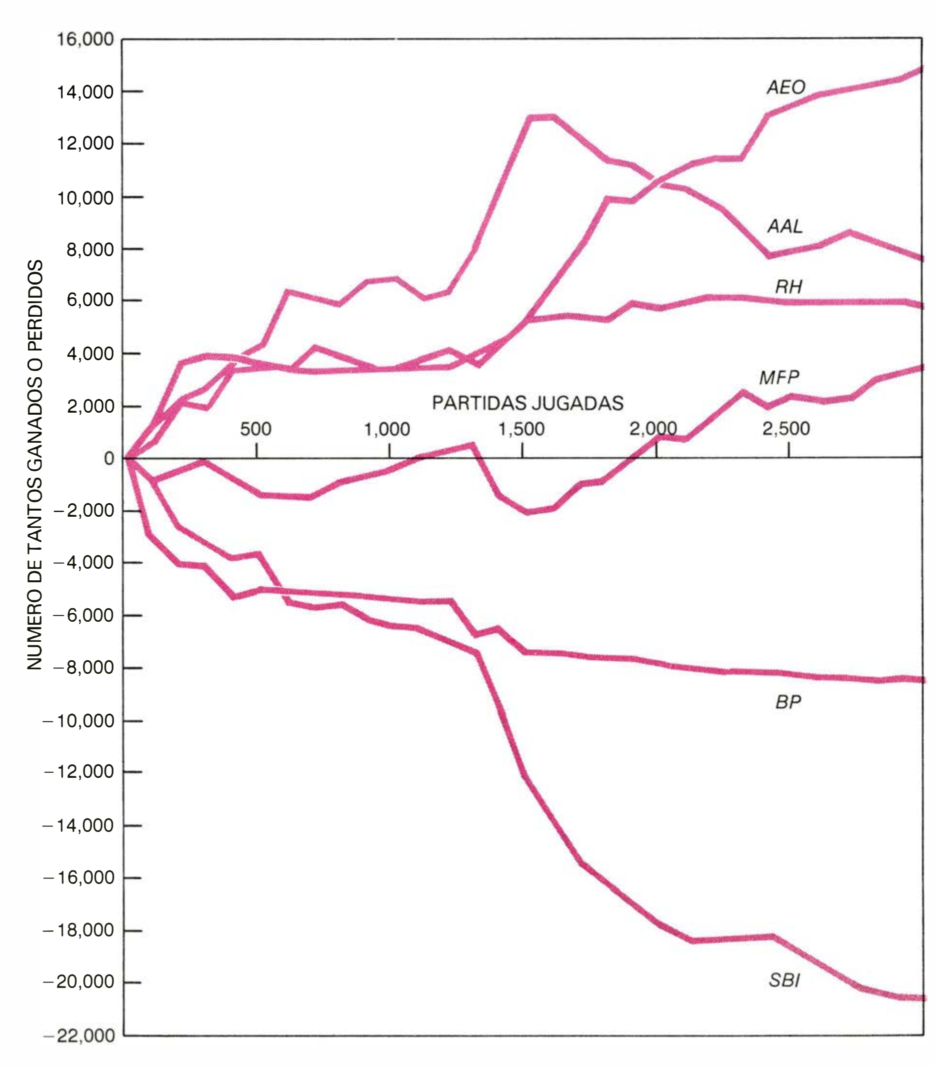


ADMINISTRACION A LARGO PLAZO DE FONDOS, necesaria en el póker como en otras muchas actividades humanas. El éxito de una estrategia de póker se mide, no por la cantidad de dinero que un jugador gana o pierde en una sola partida, sino por el total ganado o perdido en una larga sesión. La ilustración muestra las ganancias y pérdidas acumuladas de seis jugadores estáticos que compitieron durante 10 torneos, con un total de 3000 partidas. Los de mejores resultados han sido el Jugador Matemáticamente Equilibrado *(MPF)* y el Jugador RH *(RH)*. El jugador RH toma en cuenta, además de la fuerza de su mano, varios criterios secundarios de decisión que se combinan en una única variable llamada índice RH. Los otros cuatro jugadores $S_1 - S_4$ son estrategias estáticas basadas en árboles de decisión de distinta complejidad.

situaciones de juego, calcularon los valores del índice RH de cada situación y después consultaron obras de póker escritas por expertos con el fin de determinar la acción más recomendable en cada caso. Además, fue posible establecer una correlación entre los valores del índice y las acciones a seguir. De hecho, Rachlin e Higgins consiguieron descomponer el intervalo de valores del índice en segmentos, para cada uno de los cuales, según la fuerza de la mano del jugador, hay una única acción recomendada. Dicha partición ha demostrado ser muy satisfactoria. El Jugador RH obtiene muy buenos resultados al enfrentarse a otros también estáticos, incluido entre ellos el Jugador Matemáticamente Equilibrado.

La parte más interesante de la investigación que mis discípulos y yo estamos realizando no la forman los jugadores estáticos. En nuestro proyecto estamos concentrando los esfuerzos para desarrollar jugadores que aprendan: programas que elevan su nivel de rendimiento y van perfeccionando su juego conforme adquieren experiencia. En el momento actual, algunos de estos jugadores con aprendizaje apenas si son algo más que un componente de aprendizaje ligado a un conjunto básico de reglas de juego. Estos jugadores todavía no pueden ser considerados estrategias independientes y competitivas, como ocurre con los jugadores estáticos.

Se están desarrollando torneos de póker con estos jugadores de aprendizaje, al ob-



LOS JUGADORES DE APRENDIZAJE van mejorando su propia estrategia conforme acumulan experiencia. El gráfico muestra las ganancias y pérdidas acumuladas a lo largo de 3000 partidas de cuatro jugadores de aprendizaje y dos jugadores estáticos. (La inclusión de los jugadores estáticos tiene la finalidad de dar marcas de referencia para evaluar la velocidad de mejora de las estrategias y, también, de servir como puntos de contraste de los procesos de aprendizaje. El jugador con aprendizaje por Evaluación Adaptativa de Oponentes *(AEO)* efectúa una primera evaluación de la fuerza del juego de los contrarios, refinándola conforme avanza el juego. El Jugador que Vende y Compra Imágenes *(SBI)* hace faroles lo mismo que un jugador humano a fin de proyectar ("vender") una imagen de sí mismo que despiste a sus contrarios y de identificar ("comprar") la imagen, o estilo de juego, de sus adversarios. El Jugador Bayesiano *(BP)* es capaz de realizar inferencias inductivas, y aprende a mejorar su juego comparando los resultados previstos de sus acciones con las reales a lo largo de una serie de partidas. El Jugador de Nivel de Aspiración Adaptativo *(AAL)* reproduce un mecanismo que se cree existe en la cognición humana: un nivel de aspiración flexible que decide hasta qué punto vale la pena correr un riesgo teniendo en cuenta el status del jugador y su historial. El nivel de aspiración de un programa se reajusta ante cambios inesperados en la situación de juego.

jeto de determinar cómo funcionan los elementos de aprendizaje al someterlos a distintas condiciones, cómo interaccionan entre sí y cómo puede modificarse la velocidad con que mejoran su rendimiento. En estos experimentos se incluyen también jugadores estáticos con la finalidad de reducir la complejidad del proceso de evaluación de resultados cuando los programas de aprendizaje solamente aprenden de otros programas del mismo tipo.

Se ha diseñado cierto número de componentes de aprendizaje, destinados a realizar distintas funciones; actualmente están siendo ensayados. Por ejemplo, el jugador llamado de Evaluación Adaptativa de Oponentes comienza efectuando una estimación grosera de la situación de cada uno de sus contrarios, y después se sirve de la experiencia acumulada sobre el estilo de juego del oponente para mejorar continuamente la estimación. Esta estimación toma la forma concreta de una lista de posibles fuerzas de la mano del contrario, clasificadas según su plausibilidad. Conforme el jugador va recogiendo información relativa al estilo de juego de los demás participantes la lista de posibilidades se va refinando, siendo progresivamente más fiable. El proceso de adiestramiento tiene lugar a lo largo de tres dimensiones del programa: en la compilación estadística de las actividades de cada uno de los jugadores restantes, en la selección de las combinaciones de sucesos a observar en cada oponente con el fin de obtener la máxima información referente a él y en el refinamiento de la lógica de las conjeturas que en el pasado se hicieron respecto de la mano de cada adversario.

El Jugador de Nivel de Aspiración Adaptativo se fundamenta en la creencia de que, en la cognición humana, las decisiones de adopción de nuevas líneas de acción se activan por intermedio de un mecanismo que podríamos denominar "nivel de aspiración". El nivel de aspiración de un momento dado puede considerarse resultado de factores que tienden a equilibrarlo, como, por ejemplo, el costo de perseguir mejores resultados y, por otra, el valor de satisfacción con el actual grado de éxito. Los niveles de aspiración de los humanos pueden subir o bajar, respondiendo a los cambios que se perciban en la configuración de sucesos y de sus valores estimados. Por ejemplo, un jugador de temperamento conservador que pierda repentinamente una fuerte suma de dinero podría comenzar a jugar osadamente, porque desearía con ahínco enjugar sus pérdidas. El componente de aprendizaje del programa del jugador contiene un dispositivo de nivel de aspiración, que va ajustándose según que, pongamos por ejemplo, el status financiero del jugador cambie mucho más de lo esperado o cuando la probabilidad atribuida a ciertos sucesos se modifica de forma importante. Este jugador es bastante hábil en la elección de compromisos entre conducta defensiva y conducta ofensiva, recuperación de pérdidas y protección de ganancias, actuaciones tendentes a evitar fracasos y actuaciones encaminadas a buscar éxitos, etcétera.

Los dos jugadores de aprendizaje siguientes reproducen los tipos de farol característicos del juego entre humanos. El Jugador que Vende y Compra Imágenes practica el faroleo con objeto de impartir información acerca de sí mismo que despiste a sus contrarios y obtener información respecto de la verdadera personalidad de sus adversarios. Lo mismo que un jugador humano, este programa vende una imagen a base de adherirse durante un largo período a un determinado estilo de juego, y actuando después en contradicción con tal estilo, a fin de aprovechar la desorientación de algún contrario para obtener beneficios que embolsarse a su costa. El programa compra una imagen al establecer una relación entre las acciones de un contrario y las situaciones que ese adversario ha afrontado en el juego.

El componente de aprendizaje de este jugador está construido en torno a una métrica que caracteriza la "distancia" entre la imagen a vender o comprar y la estrategia matemáticamente equilibrada. Este jugador toma también en cuenta elementos de la personalidad de los adversarios. A grandes rasgos podría decirse que cuanto mayor sea la distancia entre la imagen a vender y la del Jugador Matemáticamente Equilibrado, y cuanto más cautos parezcan ser los contrarios, tanto más tiempo será necesario para cerrar la transacción de venta de imagen. Análogamente, cuanto más extremado sea el estilo de juego de un contrario, tantas más partidas habrán de jugarse antes de que nuestro jugador se decida a "comprar" (aceptar como auténtico el estilo de juego del adversario).

El Jugador Estadísticamente Equilibrado hace faroles de manera distinta. Corrige las insuficiencias del Jugador Matemáticamente Equilibrado mediante un componente de aprendizaje que identifica los faroles de sus oponentes, juzga sobre la amplitud y frecuencia de los faroles, y adapta su estilo de juego en consecuencia. También puede reconocer estrategias sin faroleo, como la del jugador Matemáticamente Equilibrado, y entonces, farolear lo necesario para vencerlas. El componente de aprendizaje del Jugador Estadísticamente Equilibrado es muy flexible y puede ajustar su estilo de juego para amoldarse a jugadores con todo tipo de perso-

nalidades, desde los más extravagantes a los más tímidos. Por consiguiente, se comporta con éxito contra otros jugadores programados.

Se han desarrollado otras cuatro estrategias más, todas de aprendizaje, capaces de efectuar inferencias inductivas, es decir, de extraer conclusiones generales a partir de sucesos específicos. Las cuatro estrategias deducen conclusiones y formulan generalizaciones a partir de ellas, examinando acciones de juego que han demostrado tener éxito en situaciones determinadas. Se diferencian entre sí por la cantidad y el tipo de informaciones que recopilan. Por consiguiente, la velocidad de mejora de estrategias es diferente en las cuatro e igual ocurre con su capacidad de juego.

Estas estrategias se denominan estrategias bayesianas, en razón del teorema de Bayes sobre probabilidad; van ajustando continuamente sus propias reglas de decisión, comparando los resultados previstos de los sucesos con sus resultados reales. Fundamentalmente hay tres métodos para adaptar las reglas de decisión de forma que los resultados previstos se aproximen más y más a los verdaderos. Si en las reglas heurísticas se introduce cierto número de parámetros, un componente de aprendizaje puede hacer que sus valores converjan hacia los valores casi óptimos. O bien puede hallarse experimentalmente una ordenación jerárquica óptima de las reglas heurísticas, hallando la frecuencia con que su empleo determina éxitos. Finalmente, es posible generar nuevas reglas heurísticas, ensayarlas, e incorporar las que tengan éxito a una estrategia. Este último proceder, sin embargo, es el más difícil y de más larga ejecución de los tres.

Hay cuatro programas más que están todavía en fase de desarrollo. Los describiré sucintamente, pues todos ellos tienen un amplio campo de aplicación. Por ejemplo, la Estrategia de Reconocimiento de Modelos examina juegos pasados con la intención de descubrir posibles relaciones causales entre sucesos de aquellos juegos y acciones de los adversarios. Estos sucesos pueden ser tendencias, o bien fenómenos aislados o cíclicos. (El programa consiente también que exista un lapso de tiempo entre un suceso y la acción asociada.) Las correlaciones postuladas, aunque suelen ser muy débiles, ayudan al jugador a discernir las razones ocultas de las acciones de sus contrarios, aspecto éste fundamental en póker.

El Encuestador y Seguidor de Consejos funciona de modo diferente a todos los programas que hasta ahora he descrito. El programa interactúa con un jugador humano —el usuario— utilizando un lenguaje razonablemente cercano al inglés. A este programa se le pueden enseñar estrategias francamente complicadas, en forma de principios de juego y de ejemplos particulares; todo este material es revisado por el programa desde el punto de vista de su coherencia y completitud, formulando prontamente preguntas clarificadoras y estructurando la información recibida de forma que pueda aplicarse al desarrollo real del juego.

Los dos últimos jugadores de aprendizaje en fase de desarrollo difieren de los restantes en otro aspecto, a saber, en que pretenden crear estrategias normativas: estrategias óptimas en el sentido de que dicen al jugador lo que debe hacer para hacer máximas sus ganancias a largo plazo. (Hasta el momento tan sólo he expuesto estrategias descriptivas, no-normativas.) Examinemos el jugador de tipo EPAM, que fue denominado así porque los principios que los inspiran son muy semejantes a los del Elementary Perceiver and Memorizer, proyecto diseñado por Herbert A. Simon, de la Carnegie-Mellon University y Edward A. Feigenbaum, de la Universidad de Stanford, y que se propone crear modelos de conducta de aprendizaje verbal. Este jugador construye un árbol de decisión descriptivo para cada adversario, que caracteriza el método que dicho adversario utiliza para discriminar las distintas situaciones del juego y decidir la acción oportuna. El jugador de tipo EPAM también construye otro árbol para el conjunto completo de sus adversarios: un árbol que muestra la forma óptima de actuación propia en todas las posibles situaciones del juego. Los árboles de los dos tipos van creciendo a medida que es necesario, siendo modificados por pequeños pasos continuos y también a saltos bruscos, simulando así la manera de aprender de los humanos.

El Cuasi-Optimizador busca generar una teoría normativa para jugar al póker, analizando diversas estrategias de máquina de carácter descriptivo (teorías descriptivas del juego), combinando después distintos componentes extraídos de ellas con el fin de construir un "super jugador". La actuación del Cuasi-Optimizador se descompone en tres fases. En la primera, el jugador acepta cierto número de estrategias de máquina (a modo de programas de "caja negra" impenetrable) y construye un modelo de cada una de ellas, por ejemplo, mediante un árbol de decisión. En la segunda etapa se asignan créditos: se juegan con estas estrategias cierto número de partidas para determinar cuáles de sus componentes son responsables de buenas consecuencias y cuáles culpables de las malas. En la tercera fase se sintetiza un nuevo programa de juego, seleccionando

de las estrategias de máquina disponibles el mejor elemento para cada segmento distinguible en la nueva estrategia. Se dice que esta nueva estrategia es casi óptima porque solamente es óptima al jugar contra el conjunto particular de estrategias oponentes utilizadas durante la fase de experimentación. Para otro conjunto de nuevos oponentes es preciso generar una nueva estrategia óptima.

Evidentemente, las estrategias de aprendizaje, como el Evaluador Adaptativo de Adversarios, el Encuestador y Seguidor de Consejos y otras tiene numerosas aplicaciones ajenas por completo al póker. En general, los programas de aprendizaje son importantes porque señalan las futuras direcciones de la programación de computadores. En último término, sería posible que los programadores tuvieran que realizar menores esfuerzos para perfeccionar sus programas, y que individuos de menor habilidad de programación se comunicaran con los computadores.

Se pueden utilizar computadores para resolver problemas en los que ni existen soluciones algorítmicas (donde cada paso de la solución está cuidadosamente especificado), o para los que dichas soluciones serían imposibles en la práctica. Como ha demostrado la investigación que acabo de describir, un programa de computador puede reproducir los métodos humanos de llegar a una decisión. El hecho de que los programas reproduzcan realmente acciones humanas indica que las decisiones que adoptan se basan en los mismos motivos que los seres humanos utilizan en sus procesos cognoscitivos. Esta proposición está siendo corroborada por una serie de experimentos que están desarrollándose hoy en día.

En estos experimentos un jugador humano se sienta en la terminal de un computador, y se le hace jugar al póker contra cierto número de oponentes. Algunos de estos adversarios son jugadores programados y algunos otros son jugadores humanos sentados en terminales situados en otras habitaciones. El sujeto debe determinar cuáles de los, pongamos por ejemplo, siete adversarios son jugadores humanos. (Los jugadores programados están ajustados de modo que no sean delatados por la velocidad con que computan sus jugadas.) Los resultados de estos experimentos todavía no pueden juzgarse concluyentes, pero parecen indicar que a menos que el sujeto conozca los mecanismos de decisión de los jugadores programados, y que él y sus adversarios humanos sean jugadores muy experimentados, el sujeto es incapaz de distinguir hombres de máquinas en un número de casos estadísticamente significativo.

Juegos matemáticos

A Charles Sanders Peirce: filósofo y experto en juegos

Martin Gardner

Podría hacerte ver toda la cuestión con tanta claridad como el sol del mediodía, si no fuera porque te has empecinado en la teoría de que no se pueden comprender las matemáticas.

—de una carta de Charles Sanders Peirce a William James

Muchos —la mayoría— de los filósofos de renombre de tiempos pasados mostraron escaso talento matemático. Pero hay algunas notables excepciones. Descartes, Leibniz, Pascal, Whitehead y Russell son tan sobresalientes en la historia de la filosofía como en la historia de la matemática. A esta reducida y selecta élite pertenece Charles Sanders Peirce (1839-1914), científico, matemático, lógico y fundador de la teoría filosófica llamada pragmatismo. En opinión de muchos, Peirce ha sido el más grande de los filósofos norteamericanos.

Peirce fue iniciado en matemáticas por su padre, Benjamín Peirce, quien en su día fue el más distinguido matemático de los Estados Unidos. No obstante, de los dos, Charles fue, con mucho, el más original de ambos. Su aportación a la lógica, los fundamentos de las matemáticas y del método científico, a la teoría de decisión y a la teoría de probabilidad fue ingente. Vale la pena destacar cuán notablemente se anticipó a desarrollos posteriores. En una época en que los infinitésimos, inventados en la infancia del Análisis, estaban completamente desprestigiados, Peirce insistía en su utilidad, punto de vista solamente vindicado con la muy reciente creación del llamado análisis No-Standard. En una época en que la física estaba totalmente dominada por el pensamiento determinista, la doctrina del "tychism", creada por Peirce, mantenía que el puro azar, que los sucesos no determinados por causas previas, es fundamental en el universo. Actualmente tal concepción es esencial en la mecánica cuántica ordinaria. Incluso la noción, expuesta por Peirce, de que las leyes naturales son "hábitos" adquiridos por un universo en desarrollo ya no parecen tan extravagantes como alguna vez sucedió. Existen respetables modelos de universos casi-periódicos en que sucesos producidos al azar crean sistemas de leyes distintos en cada oscilación. Con cada explosión, el universo desarrolla sus leyes, algunas de las cuales van cambiando conforme el cosmos envejece.

La influencia de Peirce sobre William James, amigo de siempre a quien tuvo enorme estimación, fue mucho mayor que a la inversa. La idea central de la doctrina pragmática, empezando por el título, fue presentada por Peirce en un artículo publicado en una revista popular. James recogió el término y desarrolló las ideas sugeridas por Peirce en una serie de brillantes conferencias que luego se convertirían en su libro *Pragmatismo*. Posteriormente, Peirce llegaría a irritarse tanto por las, en su opinión, temerarias exageraciones de James, que decidió transformar la voz "pragmatismo", término tan horrible, manifestaría después, que nadie se atrevería a apropiárselo.

Lo mismo que todos los matemáticos creadores, Peirce disfrutaba enormemente con el aspecto de juego intelectual que las matemáticas ofrecen. De niño se interesó vivamente por problemas de ajedrez, rompecabezas, trucos de cartas con fundamento matemático, textos en clave, etcétera. Este sentido de diversión perfunde todos los escritos de su madurez. Llegó incluso a inventar el término "musement" para describir el estado mental de especulación libre e irreprimida, aunque no tan completamente nebuloso como la ensoñación fantástica, estado en el cual la mente se traba en "puro juego" con las ideas. Tal estado mental, afirma, es la primera etapa de la invención de una buena hipótesis científica. Se medita acerca de todos los datos de importancia, y después se les va dando vueltas en la cabeza con objeto de formar nuevas combinaciones (lo mismo que se hace al ensayar movimientos de las piezas de ajedrez para resolver un problema) en espera de que llegue el misterioso chispazo de la comprensión profunda.

En un artículo casi desconocido titulado "A Neglected Argument for the Reality of God" (Un razonamiento sobre la real existencia de Dios que no ha sido tenido en cuenta) Peirce exponía que la actitud que él llamaba "musement" no solamente es una vía al teísmo, sino que, en realidad, es la única vía. Implica un salto comparable a las conjeturas de los científicos, aunque dado no por la cabeza, sino por el corazón. No es comprobable, mas para quienes lo han dado, escribía Peirce, la certeza puede ser tan grande como la seguridad en la propia existencia o en la existencia de otros. Solamente sobre temas de este estilo, de "over belief" (sobre-creencias, en terminología de James) estaban de acuerdo Peirce y James.

Su concepción de las matemáticas como elemento recreativo resulta particularmente evidente en su opinión de cómo debe enseñarse esa disciplina a los niños. Estaba convencido de que los métodos entonces utilizados solamente podrían producir azotes. Los manuscritos de sus tres libros de texto —no publicados— están llenos de nuevo modos de utilizar rompecabezas, juegos y juguetes para presentar nociones matemáticas. Por ejemplo, las paradojas de Zenón le llevan a examinar las nociones de continuidad y límite. La geometría proyectiva y las sombras producidas por una rueda giratoria iluminada por una lámpara sirven para presentar la noción de infinito. Peirce se dio cuenta —¡antes de 1900!— del gran valor de la topología elemental para estimular la imaginación matemática de los niños. (Decía que la topología era "la más sencilla, más elemental y más fundamental rama de la geometría".) La fórmula de Euler para esqueletos de poliedros, la teoría de nudos, la teoría de grafos, la conjetura del mapa de cuatro colores (que Peirce se esforzó vanamente durante décadas en demostrar), la banda de Moebius —he aquí tan sólo algunas cuestiones de la topología que Peirce utilizó para excitar el interés de los estudiantes. Le encantaba pedirle a maestros y profesores que le permitieran instruir grupos de chavales que detestasen las matemáticas y que parecieran incapa-

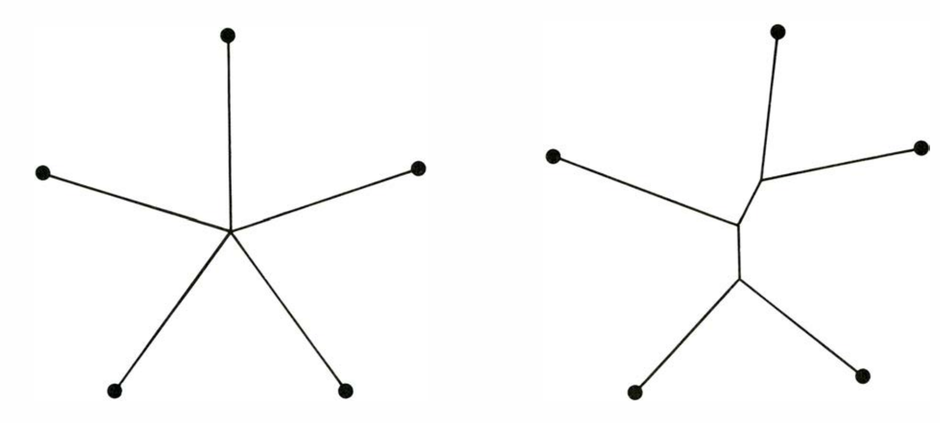
ces de aprenderlas. El mismo anotó en una ocasión que, después de unas diez lecciones, dos de sus "tontos de remate" llevaban la voz cantante en la escuela.

Para la enseñanza de la aritmética Peirce recomendaba utilizar cuentas –por ejemplo, garbanzos– siempre que fuera posible, introducir enseguida la notación binaria, usar una colección de 101 tarjetas con los números de 0 a 100, y otros recursos hoy habituales en la escuela elemental. En un libro de texto quiso insertar un dispositivo mecánico, en cartulina, para realizar multiplicaciones. "La objeción que podría presentar su inserción", anotó en un cuaderno, "es que los maestros no comprenderían los principios matemáticos en que descansa su funcionamiento, viéndose expuestos por tanto a preguntas comprometidas".

También recomienda utilizar naipes. "Si logras hacerte, querida Bárbara, con un mazo completo de naipes, comodín incluido, o sea, 53 cartas en total, te haré tragar una leccioncita de matemáticas tan fácilmente como una cucharada de aceite de ricino con un vaso de leche". Bárbara es un personaje de uno de los libros de texto, inéditos, de Peirce. La razón de llamarse así está en que "Barbara" era el nombre mnemónico medieval del silogismo "Todo *A* es *B*, todo *B* es *C*; por consiguiente, todo *A* es *C*".

En la introducción de otro de sus textos, Peirce dedica nada menos que 15 páginas al juego del tatetí (también llamado "tres en raya"). El juego sirve para mostrar cómo se empieza por conjeturar un teorema, que a continuación se comprueba manipulando diagramas. "Tales son las herramientas con que trabaja el matemático", escribió. No ha sido hasta muy recientemente, a consecuencia del enorme éxito de textos como *Mathematics, A Human Endeavor*, de Harold R. Jacobs, cuando algunos profesores se han puesto al día con las propuestas de los invencibles manuscritos de Peirce.

Lo mismo que otros muchos genios de las matemáticas –vienen a la mente Kepler y Leibniz– Peirce se volvía en ocasiones excesivamente entusiasta, casi obsesivamente, en relación con algunas de sus creaciones. En parte, ello pudo deberse a trabajar aislado, sin el toma y daca de la clase ni la constante discusión con otros colegas. Peirce no se llevaba bien con casi nadie, y en sus años tardíos, su mal carácter y su pobreza le convirtieron en un recluso solitario. James lo describió como "un pobre tipo" a quien ninguna universidad confiaría un puesto de profesor, "un ser singular" y "cascarrabias sin remedio", hombre cuyas conferencias eran "fogonazos de luz deslumbradora proyectada con la más lúgubre obscuridad". En emocionado tributo que Peirce escribió a la muerte de James, Peirce decía "¿Quién sino yo... podría ser de naturaleza más distinta de la suya? El, tan concreto, tan vital; yo, mero índice de materias, tan abstracto, retorcido como un hilo enredado".



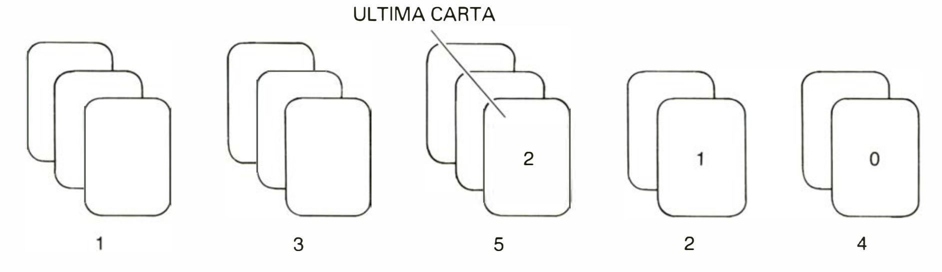
Reducción de "quinquenariedad" a "terciariedad", según Peirce

Una de las dos principales obsesiones de Peirce fue su sistema de "grafos de existencia" para representar gráficamente ideas lógicas. Al hacer notar el valor pedagógico de los diagramas de Venn para resolver silogismos o problemas más generales de álgebra de Boole, Peirce pisaba tierra firme; pero él quiso generalizar tales ayudas visuales a la lógica de todos los tipos, lógica modal incluida. Su sistema fue creciendo en complejidad, siempre basado en las propiedades topológicas del plano. Durante 20 años estuvo utilizando sus curiosos diagramas para ayudarse en sus razonamientos, y no cabe duda de que le fueron enormemente útiles. Los calificó de ser *chef d'oeuvre*, y estaba convencido de que "enseñándolos a niños y niñas antes que la gramática... les servirían de ayuda toda la vida". Si los especialistas de lógica adoptaran su método, escribió, "pronto, la lógica experimentaría tal avance que todas las ciencias acusarían sus beneficios". Desdichadamente, nadie más encontró útiles sus diagramas, aunque quizá sea demasiado pronto para emitir veredicto final. Una excelente monografía, original de Don D. Roberts, titulada *The Existential Graphs of Charles S. Peirce* fue publicada en 1972.

Y la otra gran excentricidad de Peirce –y ahora tal vez pise terreno todavía más resbaladizo– fue su convicción de que en todas las ramas de la filosofía, la organización conceptual más eficiente se obtendría mediante tres categorías fundamentales que él denominó "firstness", "secondness" y "thirdness", que podrían traducirse por primariedad, secundariedad y terciariedad. Lo mismo que les ocurre a los científicos, los filósofos se ven en la necesidad de clasificar ideas, y como la filosofía se interesa por todas las formas de conocimiento, sus esquemas suelen incluir una lista de categorías consideradas más fundamentales. Las diez categorías de Aristóteles han tenido tan enorme influencia en la filosofía occidental que, hasta que Kant propuso un sistema diferente, nunca tuvo el esquema aristotélico seria competencia. Kant introdujo doce categorías (agrupadas en cuatro tríadas) que él consideró esenciales para describir la forma en que el pensamiento consciente del hombre impone organización sobre el vasto, y en última instancia, incognoscible, océano del ser. Desde Kant hasta hoy se han creado tantos sistemas distintos que actualmente la noción de "categoría" se ha convertido en un concepto desdibujado y pasado de moda.

Peirce estaba firmemente convencido de que la más útil herramienta del instrumental filosófico era la ordenación de las cosas en mónadas, díadas y tríadas. La primariedad considera cada cosa como un todo en sí misma, como por ejemplo, la cualidad de rojo, que para entendernos podríamos denominar "rojedad". No se trata de un objeto rojo, sino de la pura posibilidad de lo rojo: "La rojedad incluso antes de que en el universo algo fuese rojo". La secundariedad considera una cosa en relación con otra, por ejemplo, una manzana roja. Se trata de la rojedad ligada a una manzana, un "hecho en bruto" del mundo real. La terciariedad se refiere a dos cosas puestas en relación ("mediadas") por una tercera, por ejemplo, una manzana que cae de un árbol. El árbol y la manzana están ligados por la relación "caer de". Nuestra representación mental de "manzana roja" es otra terciariedad, porque hace intervenir manzana, rojo y mente. El universo, "lo de afuera", mudable con el tiempo, y el mundo interior de la propia consciencia, son otros tantos dominios "reales" de terciariedad.

Peirce aplicó las nociones de prima, se-



Recomposición de cinco pilas del juego de Peirce

cunda y terciariedad a todas las ramas de la filosofía. No es preciso, arguyó, continuar con "cuaternariedad", "quinquenariedad" y sucesivas, porque en casi todos los casos estas relaciones de orden superior pueden reducirse a combinaciones de las tres primeras. Por otra parte, la terciariedad genuina es tan imposible de reducir a secundariedad como es imposible de reducir a primariedad la auténtica secundariedad. Con un hábil recurso de teoría de grafos, Peirce consiguió un modelo que exponía estas relaciones. Representemos la primariedad mediante un punto y la secundariedad mediante los extremos de un segmento rectilíneo. La terciariedad se representa mediante los extremos de tres segmentos que se unen en un punto común, como si fuera el mapa de una bifurcación de carreteras. ¿Por qué no proseguir con cuatro, cinco, seis, etcétera, haciendo confluir en un mismo punto más líneas? Porque tales "estrellas" de orden superior pueden siempre reducirse a terciariedad sustituyendo el punto central por gráficos triádicos, como se ve en la ilustración de la página anterior. Sin embargo, no hay modo alguno de conseguir la reducción de un gráfico triádico a otro con dos puntos extremos.

Peirce consideraba que sus tres categorías eran su máxima aportación a la filosofía. Negó siempre la acusación de estar enamorado del número tres. Admitía tener la hegeliana "querencia por el 3", pero subrayaba que ello se explicaba por las innumerables aplicaciones de la terciariedad. En su juventud, decía, hubiera opinado que sus categorías eran "chifladuras", pero su estudio le había convencido de lo contrario.

"El hecho más fundamental referente al número tres", escribió Peirce, "está en su potencia generativa,... Tan prolífica en formas es la tríada que podría fácilmente concebirse que toda la variedad y multiplicidad del universo emana de ella". De su grafo para representar la tríada llegó a decir que era "emblema de fecundidad, comparado con el cual el falo sagrado de la infancia de las religiones no pasa de ser sino un simple palo".

No es este el lugar de discutir la fertilidad de las categorías de Peirce, por lo que me contentaré con decir que en tiempos de Peirce el único filósofo eminente que compartió su entusiasmo por uno, dos y tres fue Josiah Royce. James se quejó de no haber podido comprender nunca las categorías de Peirce. Entre los filósofos de hoy en día de renombre mencionaré los dos únicos que conozco, Eugene Freeman y Charles Hartshorne. "Estoy convencido", escribe Hartshorne, "que todas las cosas, desde los átomos hasta Dios, son realmente manifestaciones de Primero, Segundo y Tercero, y de que ninguna otra doctrina de la misma sencillez tiene la potencia y precisión que ésta posee".

Dediquemos ahora nuestra atención a cosas menos controvertidas: trucos de cartas. En el número de abril de 1908 de *The Monist* se publica un artículo de Peirce dedicado a "Some Amazing Mazes" (Algunos sorprendentes misterios) que comenzaba con una oportuna descripción de una "danza mística" de los ángeles, tomada del *Paraíso perdido* de Milton (Libro V, 623-624):

> ...Mazes intrincate,
> Eccentric, intervolv'd, yet regular
> Then most, when most irregular
> they seem.

"Hacia 1860", comienza diciendo Peirce, "puse a punto una *mélange* de efectos de la mayoría de los principios elementales de la aritmética cíclica; y desde entonces, al final de alguna velada de jugar a las cartas, la he ido exhibiendo ocasionalmente, presentándola como "truco"... con el resultado sistemático de interesar y sorprender a los presentes, aunque sus capacidades para las matemáticas hayan ido desde la justamente suficiente para tolerar altruistamente los naipes hasta la de algunos de los más potentes matemáticos de entonces, quienes con certeza hubieran podido desvelar el misterio sin más que un poco de reflexión".

Lo que Peirce llamaba aritmética cíclica se llama hoy aritmética de congruencias. Algunos profesores la llaman "aritmética del reloj", pues éste proporciona un excelente modelo para comprenderla. Por ejemplo, 2 es igual a 14 modulo 12. Ello significa que al dividir 2 y 14 entre 12 (que es el módulo) el resto es, en ambos casos, 2. En relación con el reloj, transcurridas 14 horas a partir de las doce del mediodía, las manecillas ocupan exactamente la misma posición que ocuparon dos horas después.

El primero de los trucos de Peirce, que se publica en el Volumen 4 de sus *Collected Papers*, con el título de "The First Curiosity", es, con seguridad, el más fantástico y complicado truco de cartas que nunca se haya inventado. No puedo recomendarlo como medio de entretener a un grupo de amigos a menos que éstos sean apasionados de la teoría de números, pero en cambio, para un profesor que desee "motivar" a sus alumnos al estudio de las congruencias, es un medio soberbio. Es imposible demostrar que el truco funciona siempre sin antes aprender gran cantidad de "aritmética cíclica", incluido un famoso teorema de Fermat sobre números primos.

Antes de proseguir leyendo, se insta al lector a hacerse con una baraja francesa (de 52 naipes) y a seguir escrupulosamente el procedimiento. Se separa todos los corazones y se los coloca en el orden sucesivo, de as a rey, formando un paquete que colocado boca abajo deje al as en lo alto. Se hace lo mismo con las espadas, aunque sin el rey, que no se usa. Así pues, el paquete de espadas está formado por 12 cartas boca abajo, con el as en lo más alto y la reina en lo más bajo. Se coloca el paquete rojo boca abajo sobre la mesa, y se sostiene el paquete negro, también boca abajo, en una mano.

Se van dando las cartas negras, colocándolas en dos pilas boca arriba. (Siempre que se repartan cartas para colocarlas en dos pilas, se las sostiene boca abajo y se colocan boca arriba, de izquierda a derecha, comenzando por la izquierda.) La última carta a repartir (la reina) se desecha, colocándola aparte, donde se formará una pila de cartas desechadas. En su lugar se coloca la carta superior (en este caso el as) del paquete de cartas rojas, colocándolo boca arriba en la segunda pila, en lugar de la reina recién desechada. Se reúnen las dos pilas tomando la situada más a la izquierda y colocándola boca arriba sobre la otra pila. Se vuelve el paquete negro (que ahora contiene una carta roja) boca abajo y se repite exactamente el mismo procedimiento. Esta vez será el dos rojo quien reemplace a la última carta negra (que es la sota).

La sota negra se coloca boca arriba sobre la reina negra anteriormente desechada. Se reitera el procedimiento hasta realizarlo en total 12 veces. Quizá le sorprenda descubrir que ahora se tiene un

paquete formado totalmente por cartas rojas y que la pila de desechos contiene todas las cartas negras. Se recoge el rey de corazones restante y se coloca bajo el montón de cartas rojas boca abajo.

Para asegurarse de que todo el procedimiento se ha seguido correctamente, compruébese el paquete rojo. Manteniéndolo boca abajo, y leyendo desde arriba, el orden de las cartas debe ser: 7,8 *J*, 9,4 *Q*, 6, 10, 3, 5, 2, *A*, *K*. El paquete negro ha de ser: *Q*, *J*, 9, 5, 10, 7, *A*, 2, 4, 8, 3, 6.

Existe una curiosa correlación entre ambos paquetes. El valor de la carta situada en la *n*-ésima posición contada desde arriba en cualquiera de los dos paquetes da la posición, contando desde arriba, que en el otro paquete ocupa la carta de valor *n*. ¿Por ejemplo, dónde está la sota de espadas? Dándole a la sota el valor de 11, se mira la undécima carta del paquete rojo. Es un 2. Tomemos la segunda carta del paquete negro. Es la sota de espadas. ¿Dónde se encuentra el cinco de corazones? La quinta carta del paquete negro nos lo dirá. Es un 10. Pues bien, la décima carta del paquete rojo es el cinco de corazones.

Antes de revelar a su auditorio esta notable correlación, conviene simular una barajadura, siguiendo el procedimiento que a continuación se expone. Primero se deja que alguien corte el paquete de cartas rojas en dos, tantas veces como se quiera. Manteniéndolo boca abajo, pídasele a alguien presente que diga un número, de 1 a 12. Llamemos *k* a este número. Entonces el paquete se reparte en *k* pilas de cartas boca arriba; a continuación, se reconstruye el paquete recogiendo las pilas a partir de una cualquiera, la que designe el auditorio. El paquete así reconstruido puede volver a cortarse y todo el procedimiento anterior repetirse tantas veces como se quiera, bien con el mismo valor de *k* o con otro diferente. Podría suponerse que al cortar, repartir en *k* pilas, volver a cortar y repartir, y repitiendo esta rutina con cualquier valor de *k* que se pida, mezclará sin remedio las cartas rojas. Sorprendentemente, gracias a los teoremas de la aritmética de congruencias, la correlación de los paquetes se conserva.

La única dificultad del proceso de mezclado es ésta: cuando se recomponen las *k* pilas es necesario hacerlo de una manera determinada, que depende de *k*. Pensemos en que la hilera de pilas de cartas forme un círculo, situando mentalmente la última pila junto a la primera, con lo que se puede contar "alrededor" de la hilera, sea en el sentido de las agujas de un reloj, sea en el sentido contrario. Obsérvese la pila en la que se dio la última carta. Mentalmente se llama "cero" a la última pila de la derecha y se cuenta hasta la pila que recibió la última carta. Se cuenta en el sentido de las agujas o en el contrario, el que sea más corto.

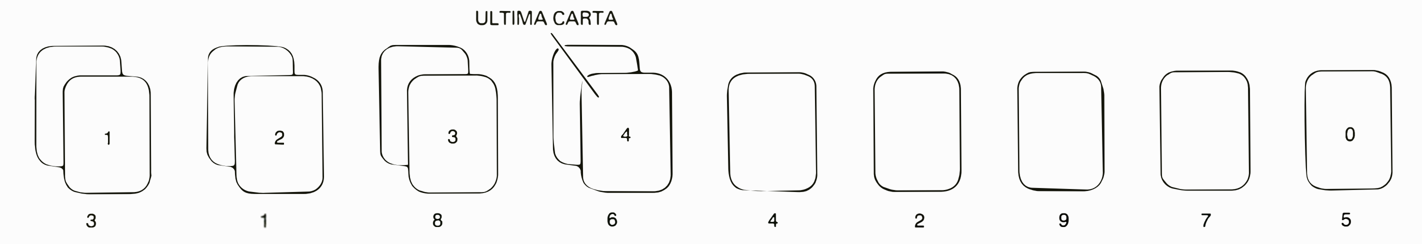
Supongamos que se hayan repartido las cartas rojas en cinco pilas de cartas boca arriba. La pila central será quien reciba la última carta, tal como queda de manifiesto en la ilustración de la página anterior. Es la segunda desde la derecha, contando en sentido horario desde el extremo derecho de la fila. Esto implica que las cartas deben recomponerse del modo siguiente. Se toma una pila cualquiera y se coloca boca arriba sobre la segunda pila situada a su izquierda, contada en sentido antihorario. Tómese la pila doble así formada y colóquesela sobre la segunda pila hacia la izquierda, prosiguiendo de este modo hasta que solamente quede un paquete. Los números situados bajo las pilas dibujadas en la ilustración muestran el orden de montaje, suponiendo que se empiece por la primera pila de la parte izquierda de la hilera.

No se debe olvidar que al recoger las pilas se están contando posiciones, no las pilas reales. Para ciertos valores de *k*, la pila que recibe la última carta es adyacente a la situada a la derecha del todo de la hilera. La recogida de pilas es entonces muy sencilla; pues las pilas a tomar siempre son adyacentes; pero si hay mayor separación, como ocurre en el caso de nueve pilas, en que la mínima distancia es cuatro, hace falta bastante experiencia para recoger rápidamente. La ilustración inferior de esta misma página muestra el orden de recogida de nueve pilas suponiendo que se empiece por la segunda desde la izquierda. En este caso se procede en sentido de las agujas del reloj, porque la cuenta más corta desde el extremo derecho de la hilera hasta la pila que recibió la última carta va en sentido horario. Para practicar conviene marcar las posiciones con una hilera de monedas. Tras cierto tiempo, éstas ya no son necesarias.

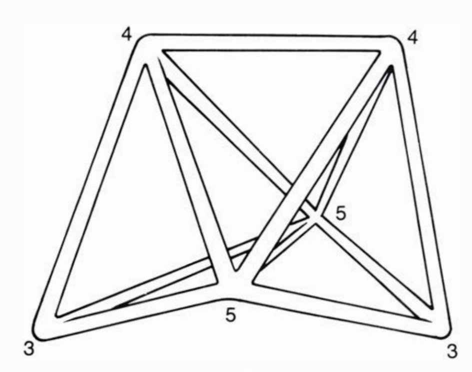
El paquete recompuesto puede cortarse siempre tantas veces como se quiera, y el "mezclado", repetirse con distinto número de pilas. Una vez que todo el mundo se dé por satisfecho con la "barajadura", es necesario darles al paquete rojo y al paquete negro un solo corte más. La pila roja se corta de manera que el rey quede situado en lo más bajo al colocar las cartas boca abajo. Peirce sugiere que se haga en ese momento la observación de que no habiendo rey de espadas, se cortará el mazo rojo de manera que el rey quede abajo, "haciendo inútil por tanto la búsqueda de dicha carta". Al actuar así préstese atención a la carta que queda en lo alto del paquete. Supongamos que sea un 4. Entonces es necesario cortar la pila negra de forma que el as quede en cuarto lugar desde arriba. ¡Ahora las dos pilas guardan la misma correlación que al principio!

Para subrayar la correlación, Peirce sugiere que las cartas negras se dispongan boca abajo en hilera. Se le pide a los espectadores que digan una carta roja. Supongamos que sea el 7 de corazones. Vaya tocando las cartas negras, una por una, contando de uno hasta siete, y vuelva boca arriba la séptima carta. Observe su valor *n*. Cuente en este momento hasta la *n*-ésima carta en el paquete rojo. ¡He ahí el 7 de corazones!

"El auditorio nunca deja de querer que se repita el experimento", escribe Peirce. Tras mostrar que cierto número de cartas se encuentran en los lugares indicados, puede repetirse el truco mezclando otra vez las cartas rojas mediante unos cuantos repartos, y reajustando después los dos paquetes mediante los cortes oportunos. "Si usted desea ahora una explicación de todo este asunto", concluye Peirce, "tal deseo muestra que no está usted muy impuesto en aritmética cíclica". Como obra de consulta para los lectores, Peirce cita un libro de Richard Dedekind, añadiendo



Recomposición de nueve pilas del juego de Peirce



El "otro" deltaedro de ocho caras

que en otra ocasión escribirá él un pequeño ensayo sobre el tema.

Peirce se "medio redime" de su "media promesa" –tal son sus palabras– en un artículo publicado en el número de julio de 1908 en *The Monist*, artículo que también figura reproducido en los *Collected Papers*.¡La explicación de Peirce ocupa nada menos que 58 páginas! El ensayo es del Peirce más puro, edición completa y no abreviada, e incluye generalizaciones, fórmulas, horrendos diagramas de existencia que más parecen obras de arte abstracto, y deliciosas digresiones sobre el valor de hacer apuntes en fichas, sobre máquinas lógicas, sobre el significado de la continuidad, el modo en que la mente influye en la materia, y la naturaleza del albedrío y del tiempo.

En 1958, Alex Elmsley, mago londinense, señaló en *Ibidem* (publicación canadiense periódica, dedicada a la magia) que en la primera fase del truco de Peirce no es esencial que sea la última carta negra la que se reemplace por una roja. La carta a sustituir puede ocupar cualquier lugar (siempre el mismo) en el paquete negro. Por tanto, se puede dejar que algún espectador elija un número n entre 1 y 12, y en cada reparto sustituir la n-ésima carta negra por una roja.

Los escritos de Peirce están actualmente recogidos en 13 volúmenes, seis editados por Hartshorne y Paul Weiss, dos por Arthur W. Burks, y cinco más por Carolyn Eisele. Es un triádico escándalo (1) que el grueso de los artículos de carácter matemático de Peirce no fuesen publicados hasta hace dos años, en que la señora Eisele recopiló hábilmente la colección para ser publicada en *The New Elements of Mathematics*, (2) que estos libros apenas hayan recibido publicidad ni hayan sido adecuadamente recensados y (3) que la obra completa cueste más de 300 dólares americanos.

La ilustración superior de esta página muestra la solución del primero de los problemas propuestos el mes pasado: hallar un deltaedro de ocho caras que no sea un octaedro regular. (Los deltaedros son poliedros cuya caras son triángulos equiláteros.) En el octaedro regular en cada vértice se reúnen cuatro caras y cuatro aristas. En el deltaedro pedido hay dos vértices que reciben tres aristas, dos más que reúnen cuatro aristas, y otros en los que se encuentran cinco aristas.

El segundo problema consistía en aplicar la fórmula de Euler con objeto de probar que no puede cubrirse una esfera con un mapa "regular" de hexágonos, en que cada vértice sea punto de reunión de tres aristas. La fórmula de Euler dice que $C - A + V = 2$. Supongamos que exista un tal mapa. Cada hexágono tiene seis vértice y seis lados. Así pues, si los hexágonos no compartieran ciertos vértices y ciertos lados, habría seis veces más aristas que caras. Sin embargo, cada vértice es compartido por tres caras; por consiguiente el número de vértices del mapa será $6C/3$. Análogamente, cada lado es compartido por dos caras; así que el número de aristas del mapa ha de ser $6C/2$. Sustituyendo estos valores en la fórmula de Euler se obtiene la ecuación $C + 6\ C/3 - 6\ C/2 = 2$, que simplificada se convierte en $C + 2C - 3C = 2$, es decir, $0 = 2$. Esta contradicción prueba que la hipótesis de existencia del mapa es falsa.

¿Qué sucede cuando el razonamiento anterior se aplicapas regulares formados por las aristas de los sólidos platónicos? Se obtiene en cada caso una fórmula que da a C, número de caras, un único valor: 4, 8 y 20, para el tetraedro, octaedro e icosaedro, respectivamente, 6 para el cubo, y 12 para el dodecaedro. Como es imposible que un poliedro regular tenga caras de más de seis lados, hemos demostrado que no pueden existir más que cinco sólidos regulares.

También la fórmula de Euler interviene en la demostración de que hay precisamente ocho deltaedros convexos. Véase "Deltahedra" en la obra *Excursions into Mathematics*, por Anatole Beck, Michael Bleicher y Donald Crowe, páginas 21 a 26.

En la sección de junio uno de los pies de figura no era enteramente correcto. El paisaje imaginado por Benoît Mandelbrot y programado por Richard F. Voss debió ser llamado superficie Browniana *modificada*. También olvidé decir que la sorprendente portada de la revista, que mostraba una curva en copo de nieve de Mandelbrot-Peano, fue trazada mediante un programa de gráficas de computador, escrito por Sigmund Handelman y Mark Laff, miembros ambos del equipo del Thomas Watson Research Center de la International Business Machines Corporation (IBM).

Taller y laboratorio

Sistemas químicos que permiten oscilar de un color dado a otro

Jearl Walker

Si se mezclan perfectamente diversas sustancias químicas, lo lógico es esperar que la mezcla adopte un color único y uniforme. Sin embargo, algunas reacciones químicas producen una sorprendente oscilación de colores. Unas veces cada minuto y otras veces de hora en hora, las soluciones cambian de color pasando de uno a otro cíclicamente.

Uno de los más famosos osciladores de este tipo fue descubierto en 1958 por el químico ruso B.P. Belousov, e investigado posteriormente por A.M. Zhabotinsky y otros. La reacción original tenía lugar al mezclar bromato potásico, sulfato cérico y ácido cítrico en ácido sulfúrico diluido. La oscilación iba del incoloro al amarillo. Desde entonces se han encontrado otras mezclas que oscilan del rojo al azul, del violeta al azul, del rojo al verde e incluso del incoloro al dorado y, posteriormente, al azul. En este trabajo se describirá cómo hacer varios osciladores de éstos y cómo investigar las causas o motivos que provocan la periodicidad de los cambios.

En este trabajo el lector podrá entretenerse con cuestiones que aún no han sido totalmente resueltas. Cabría esperar oscilaciones cromáticas en mezclas no homogéneas, ya que en ellas el motivo puede ser la posterior difusión de un ingrediente en otro. Pero las oscilaciones de un sistema cerrado consistente en una mezcla homogénea siempre se habían visto como una cosa imposible, sobre la base de que lo contrario violaría una ley fundamental de física y química, según la cual toda reacción espontánea debe hacer descender regularmente la energía libre de Gibbs del sistema (más adelante hablaremos de la energía libre de Gibbs).

En 1970 Zhabotinsky y A.N. Zaikin comunicaron que habían encontrado estructuras periódicas (ondas circulares) que recorrían de una parte a la otra una solución, cuya composición era un tanto distinta de la solución primitiva que sirvió como punto de partida. Más tarde, Arthur T. Winfree, que está ahora en la Purdue University, mejoró la fórmula con la que obtener la solución y examinó ampliamente las estructuras periódicas de las ondas espirales rotativas que aparecían en dicha solución. Winfree redactó para mí unas normas de trabajo con las que poder realizar las reacciones uno mismo, que a continuación paso a describir.

Winfree sugiere la siguiente forma de mezclar los reactivos para obtener un oscilador del rojo al azul. Añadir dos mililitros de ácido sulfúrico concentrado y cinco gramos de bromato sódico a 67 mililitros de agua pura (destilada o desionizada) con lo que se obtendrán 70 mililitros de solución. *Como norma de seguridad añadir siempre lentamente el ácido al agua y nunca el agua al ácido, ya que existe el peligro de que explote.*

Verter seis mililitros de esta solución en un recipiente de vidrio y añadirle 0,5 mililitros de una solución de bromuro sódico, obtenida agregando un gramo de bromuro sódico a 10 mililitros de agua. A continuación añadir un mililitro de solución de ácido malónico (malonato) hecha poniendo 1 gramo de ácido malónico en 10 mililitros de agua. Una vez que haya desaparecido el color del bromo poner a la solución un mililitro de solución 0,025 molar de sulfato ferroso de fenantrolina (un tinte también llamado a veces "ferroin"). Incorporar una pequeñísima cantidad (aproximadamente un gramo por litro) de, por ejemplo, Triton X-100, un modificador de la tensión superficial o "photoflo" (sustancia empleada en fotografía para los trabajos en el cuarto obscuro) con el fin de reducir la tensión superficial y, por tanto, ayudar a que el fluido pueda extenderse en una capa delgada.

Extender dicha capa de líquido en un recipiente limpio, como pudiera ser, un plato de los que se usan para preparar cultivos. Agite bien la solución y verá cómo, después de cinco minutos más o menos, el color comienza a oscilar. A partir de entonces y de vez en cuando, haga girar suavemente en el plato la solución, dándole vueltas o bien agitándola, suavemente también. Los colores se sucederán con un período de aproximadamente un minuto, aunque el azul sólo permanecerá durante unos cinco o diez segundos cada vez. También aparecerán pequeñas burbujas de dióxido de carbono en la solución: son producto de una de las reacciones involucradas en la oscilación y pueden eliminarse agitando el fluido de vez en cuando.

El hecho de que las concentraciones sean muy exactas no es fundamental, pero en cambio sí debe ponerse mucho cuidado en que los recipientes que se usen estén muy limpios y en procurar que los dedos no toquen la solución ni la superficie de los recipientes que posteriormente estarán en contacto con dicha solución. El cloruro presente en las sales de la piel humana puede entorpecer las reacciones oscilantes.

Antes de usar el bromato sódico en las soluciones, Winfree lo suele recristalizar con el fin de quitarle cualquier impureza que pudiera interferir en las oscilaciones. Dicha recristalización se hace disolviendo bromato sódico hasta la saturación en agua destilada caliente, colocada dentro de un recipiente muy limpio, teniendo cuidado de no poner directamente a la llama nada que esté empapado o bien contenga bromato sódico ya que existe el peligro de una oxidación rápida. Envolver el recipiente cerrado con un material aislante, como por ejemplo Styrofoam, y dejarlo en reposo sin moverlo durante una o dos semanas. Una vez transcurrido dicho tiempo, quítese del recipiente el líquido y, con un instrumento limpio, recójase el bromato sódico que se haya quedado pegado en las paredes y en el fondo del recipiente. El líquido sobrante guárdelo, pues le puede ser útil para otra recristalización.

Antes de examinar las posibles variaciones de la mezcla original debo explicar el por qué tienen lugar las oscilaciones de color. Las reacciones que parecen ser las responsables de dichas oscilaciones resultan un tanto complicadas. En solución, tanto el bromuro como el bromato reaccionan con el malonato para formar el

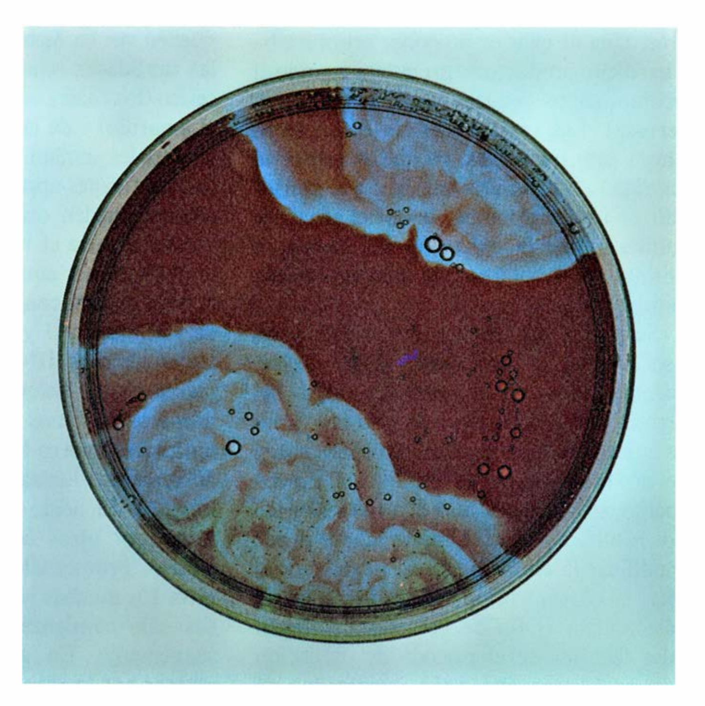
bromomalonato. Si no fuera por un efecto inhibidor del bromuro, el bromato también reaccionaría con el colorante de fenantrolina oxidando el hierro que contiene (quitando un electrón al orbital de electrones más exterior de cada átomo de hierro). El colorante en su estado reducido (ferroso) es rojo mientras que en su estado oxidado (férrico) es azul, de manera que si el bromuro dejase de inhibir la oxidación, la solución pasaría del rojo al azul.

Esto es exactamente lo que llega a suceder, ya que el bromuro es casi totalmente consumido por reacción con el malonato. Entonces puede llevarse a cabo la oxidación, el colorante de fenantrolina se oxida y la solución toma un color azul en lugar del rojo que tenía.

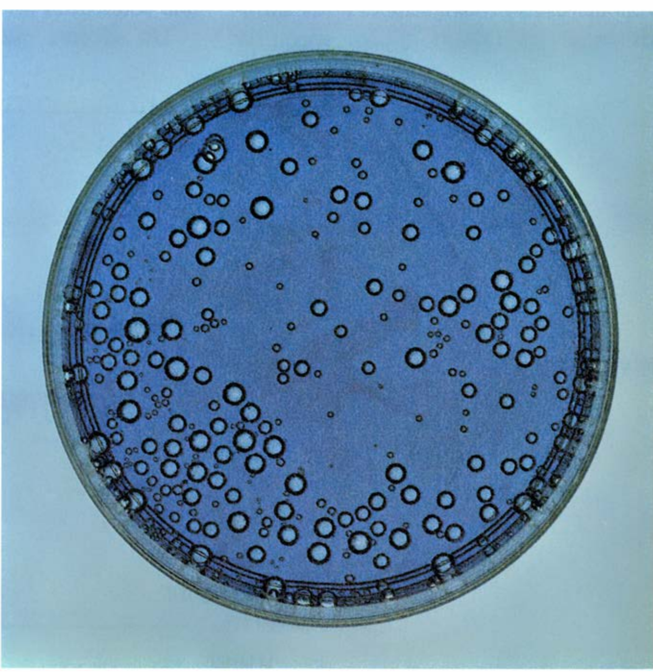
¿Por qué vuelve a tomar luego el color rojo? El bromomalonato que se ha producido y el que aún se puede producir reduce la forma férrica de nuevo a la forma ferrosa (reducción es lo opuesto a oxidación). La solución, como consecuencia, se torna de nuevo roja. Por otra parte, se produce más bromuro a partir del bromomalonato en la reacción de manera que la inhibición de la reacción de oxidación por el bromato del colorante se vuelve a producir. Rápidamente el bromuro se consume de nuevo y el ciclo se reanuda. Las oscilaciones continúan durante una hora más o menos, hasta que la solución se queda ya fija en rojo o en azul, dependiendo de las concentraciones iniciales.

Winfree ha descrito algunas modificaciones que se pueden introducir en la mezcla de partida o básica. Para ensayar con ellas hay que estar familiarizado con los términos molar y normal aplicados a las soluciones. El término molar se basa en el peso atómico de la molécula o del ion complejo que se pone en solución. En una solución uno molar el número de gramos de soluto que se ponen en un litro de









Oscilador de masa de Arthur T. Winfree modificando su color del rojo al azul

solución es igual al peso molecular del soluto (expresado en unidades de peso atómico). La concentración de un producto suele ir impresa en el recipiente que lo contiene. La fenantrolina se vende normalmente a la concentración que se requiere. Una solución normal resulta similar a una solución molar pero su concentración es además proporcional al número de carga del ion que produce la molécula al disociarse. Por ejemplo, una solución dos normal de sulfúrico es a la vez una solución uno molar, debido a que una vez puesto en solución la molécula de ácido sulfúrico libera dos átomos de hidrógeno con una carga neta de más dos.

Ahora veamos las variaciones de Winfree. Una de ellas es no poner bromuro (si bien dicho producto se producirá cuando el bromomalato reduzca a la fenantrolina ferrosa). Las oscilaciones de color serán ahora más frecuentes, ya que es menor la cantidad de bromuro que hay que consumir en cada ciclo antes de que el efecto de inhibición de la oxidación del hierro se elimine y el color de la solución pueda cambiar al azul.

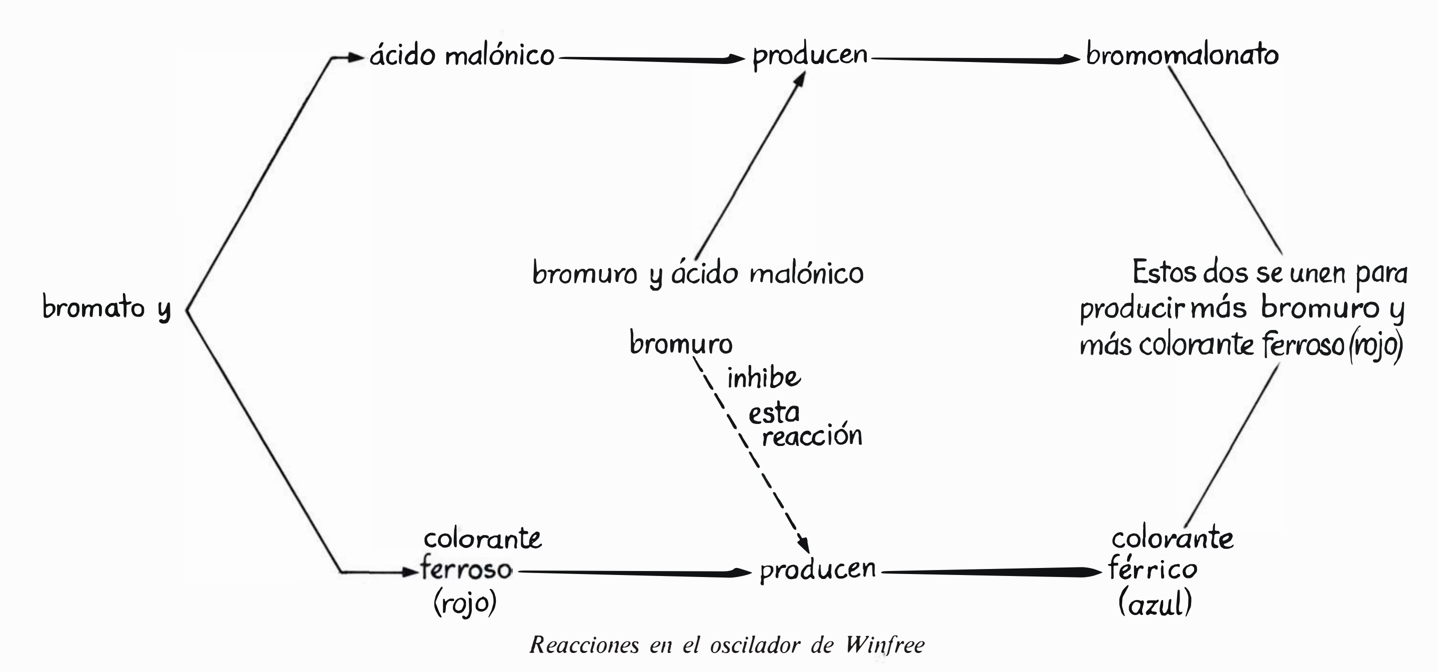
Se puede dar mayor intensidad a las oscilaciones, y así poder verlas mejor, reemplazando en la solución algo del ferroin por sulfato cérico 0,1 molar. Si no se quieren usar ácidos que manchen, por si se producen derrames durante el experimento, se puede sustituir el ácido sulfúrico por bisulfato sódico. El investigador puede modificar la temperatura de la solución o bien la concentración de los ingredientes básicos con el fin de ver cómo influyen tales factores en el período de oscilación (los profesores que estén interesados en proponer estas investigaciones a sus alumnos en el laboratorio pueden encontrar ayuda en el trabajo de John F. Lefelhocz de la Virginia Commonwealth University, cuya referencia se da en la sección de bibliografía de esta revista).

Además de las mezclas de Winfree otras pueden también producir oscilaciones de color. Las reacciones descubiertas por Belousov se basan en el sistema cérico-cerio en lugar de en el férrico-ferroso; la transición tiene lugar del incoloro al amarillo y a la inversa. Richard J. Field, de la Universidad de Montana, sugiere añadir ferroin para dar más realce a la variación del rojo al azul. Las concentraciones que él recomienda para una realización demostrativa se indican en la ilustración superior de la página siguiente. Los colores que se pueden ver en dichas mezclas dependen de las cantidades relativas de los iones cerio-cérico (los cuales oscilan entre el incoloro y el amarillo) y de las de los iones de hierro (los cuales oscilan entre el azul y el rojo) con cantidades apropiadas de amarillo; los colores pueden oscilar entre el azul y el violeta o entre el verde y el rojo.

Oscilaciones entre el amarillo-castaño y el rosa pueden conseguirse con soluciones de manganeso II y III, sustituyendo a los conjuntos cerio III-IV y hierro II-III de las reacciones de Belousov y de Winfree (los números romanos indican el estado de la ionización: hierro II es un átomo de hierro doblemente ionizado, o en el estado ferroso). El ácido malónico puede sustituirse por otros ácidos: cítrico, maleico, málico, bromomalónico y dibromomalónico. En muchas soluciones, las oscilaciones sólo comienzan cuando se las agita suavemente. En algunos casos se debe esperar por lo menos 40 minutos antes de que comiencen las oscilaciones.

Un sistema que cambie su color del incoloro al dorado luego al azul y vuelva otra vez al incoloro para repetir el ciclo puede obtenerse a un costo relativamente bajo con el reloj de yodo, descrito por Thomas S. Briggs y Warren C. Rauscher de la Galileo High School de San Francisco. Yo puse las concentraciones (en agua destilada) que ellos decían en tres recipientes distintos y las tripliqué con el fin de tener claro la cantidad de productos químicos a echar en cada uno de los recipientes. Los recipientes deberán contener todos igual cantidad, de manera que al verter el contenido de todos en un recipiente único se obtendrán las concentraciones establecidas por Briggs y Rauscher.

Poner en el primer recipiente una solución 3,6 molar de peróxido de hidrógeno, preparada añadiendo 40 mililitros de peróxido de hidrógeno al 30 por ciento a 60 mililitros de agua. *Se debe tener extremo cuidado al manejar el peróxido de hidrógeno.* En el segundo recipiente mezclar iodato potásico 0,201 molar (4,3 gramos por 100 mililitros de solución con agua) y ácido perclórico 0,159 molar preparado añadiendo a 2,3 mililitros de 70 por ciento de ácido perclórico agua suficiente como para obtener 100 mililitros de solución. Probablemente deberá calentar la solución de iodato potásico para conseguir que se disuelva completamente. En el tercer recipiente poner ácido malónico 0,150 molar (1,5 gramos por cada 100 mililitros de solución con agua), sulfato de Manganeso (II) 0,0201 molar (0,3 gr por 100 mililitros de solución con agua) y 0,03 por ciento de almidón. (Los porcentajes se refieren al porcentaje en peso del componente presente en la solución. Es decir, 0,3 gr de almidón en un litro de agua destilada



Reacciones en el oscilador de Winfree

equivale a una solución con el 0,03 por ciento de almidón.)

Cuando lo tenga todo preparado para observar las oscilaciones mezcle cantidades iguales de cada una de las tres soluciones. El azul se presentará debido al complejo azul de almidón que se irá formando periódicamente conforme la concentración de ioduro se aproxime a su valor máximo. Briggs y Rauscher obtenían oscilaciones más cortas cuando sustituían el ácido malónico con 2,4 pentadiona. Las oscilaciones son más rápidas si se reemplaza el manganeso por cerio.

Para obtener grandes oscilaciones de color en la mayoría de las mezclas la solución debe ser agitada de forma continua (con un agitador magnético si se tiene acceso a alguno) o bien se ha de hacer girar el recipiente también de forma continua. Además de los anteriores, también tiene lugar en estas soluciones otros fenómenos de entre los cuales uno de los más interesantes es el que ha sido estudiado a fondo por Winfree y que son las ondas de color que aparecen sobre y en la solución, las cuales se propagan a su través a una velocidad de unos cuantos milímetros por minuto, girando alrededor de su punto de origen, destruyéndose al chocar una con otra y describiendo una gran variedad de formas: anillos, elipses y espirales.

Para dar un mayor realce a estas ondas de color (disminuyendo las oscilaciones en masa) Winfree sugiere usar más bromuro y menos ácido sulfúrico en la mezcla que sirve de base. Por ejemplo, en la solución de bromato sódico usar un mililitro de ácido sulfúrico concentrado y 67 mililitros de agua destilada.

Para que las ondas se puedan ver mejor, poner un poco de la solución en un recipiente que tenga como un milímetro de fondo, un plato por ejemplo, para preparar cultivos de plástico. El recipiente debe estar muy limpio ya que cualquier pequeña mota de suciedad o raspadura pueden provocar la formación de burbujas de dióxido de carbono durante el experimento y, como consecuencia, generar demasiadas ondas a la vez. El contraste visual se mejora si se sitúa el plato sobre otro plato que contenga sulfato de cobre azul con unas cuantas gotas de ácido sulfúrico; el conjunto de los dos platos se coloca sobre una luz. La solución del plato inferior absorberá el calor emitido por la luz impidiendo así que dicho calor interrumpa o destruya las ondas que se forman en la solución del plato superior. El plato se debe cubrir con su tapa, para así proteger las ondas de las corrientes de aire y eliminar además la evaporación. Deje el plato en reposo sin que nada lo mueva. Si las ondas no comienzan a formarse a los pocos minutos, toque la solución con una aguja caliente, ésta las hará aparecer. Normalmente habrá en la solución burbujas o motas de polvo capaces de hacer aparecer algunas ondas.

Reactivos	Rangos de concentraciones (molar):	Concentración (molar) adecuada:
Acido malónico	0,0125 — 0,5	0,275
Nitrato de amonio y cerio	0,0001—0,01	0,002
Acido sulfúrico	0,5 — 2,5	1,5
"Ferroin"	0,0006	0,0006
Bromato potásico	0,03—0,0625	0,0625

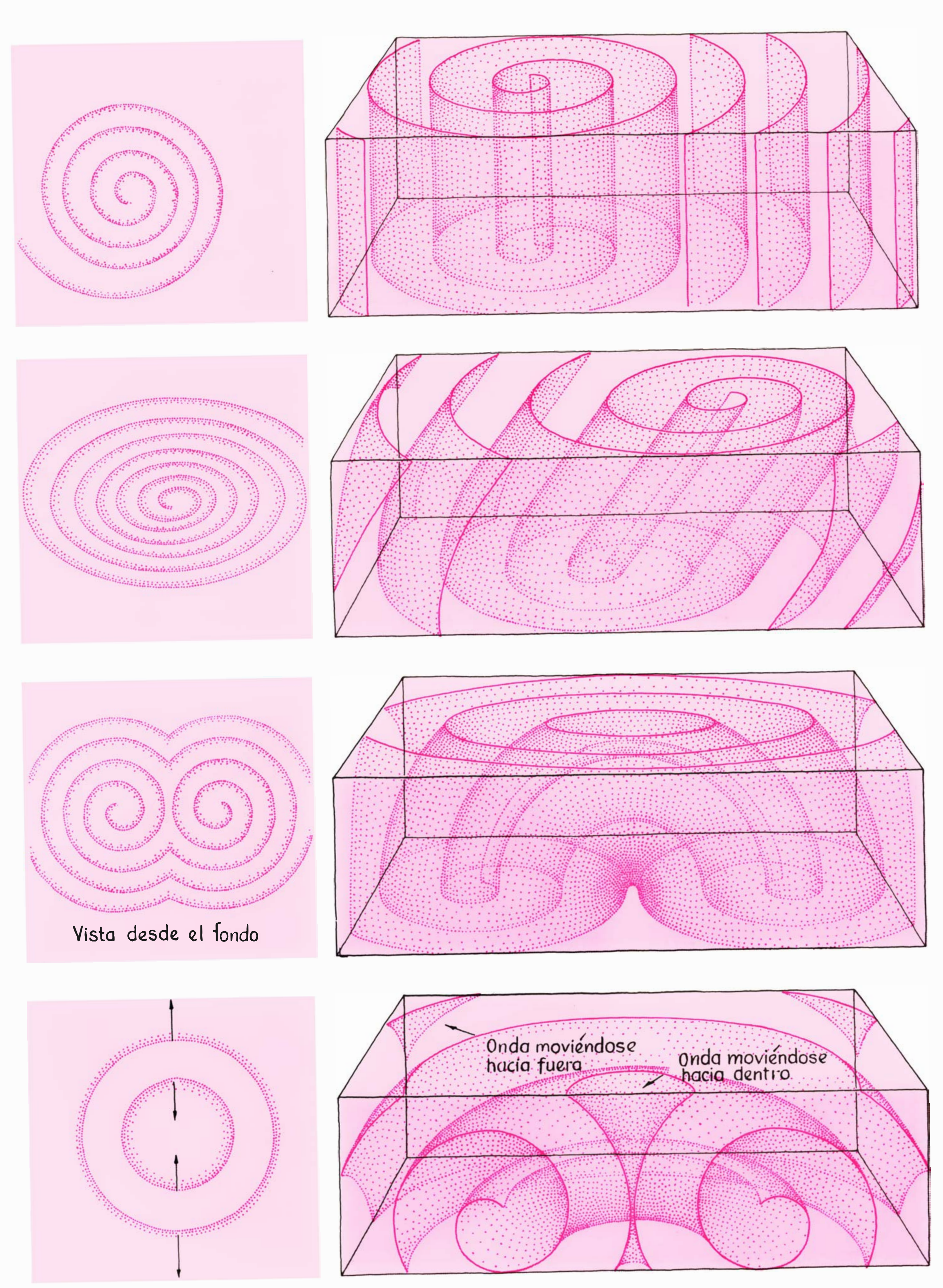
Concentraciones de Richard J. Field para el oscilador de cerio

Es de suponer que las ondas provienen del mismo proceso químico que provoca las oscilaciones de color de toda la masa. La solución permanece roja con alguna concentración de equilibrio de bromuro. Una vez que la mezcla ha sido alterada por una burbuja, una raspadura, una mota de polvo o una aguja caliente, parte de la solución consume su bromuro. El bromato puede entonces cambiar el color de la solución a azul oxidando el hierro de la fenantrolina. Esta conversión se propaga hacia fuera como una onda ya que a medida que la difusión transporta bromuro de la zona roja inmediatamente exterior al anillo azul a la zona azul, se consume el bromuro adicional y la zona antes roja se vuelve azul. Una vez que el anillo azul ha pasado una zona determinada, la solución vuelve a recuperar su tono rojo a medida que el bromomalonato reduce el hierro de la fenantrolina y desprende más bromuro.

Las ondas son realmente químicas más que hidrodinámicas, ya que el fluido no se mueve. Más aún, no reflejaría las barreras colocadas a su paso y las ondas movién-



Modelos de onda en espiral en una mezcla de Winfree sin agitar



Tipos de ondas espirales

dose en direcciones opuestas no pasarían unas a través de las otras.

El experimentador podrá comprobar que las ondas, que tienen una longitud de unos cuantos milímetros, toman distintas formas. Entre ellas encontramos las de anillos cerrados, concéntricos formando un ojo de buey, con la aparición de nuevos anillos en el centro a intervalos regulares; estos modelos no giran. Otras ondas forman espirales, cada una de las cuales tienen el mismo período de duración, más corto que ninguno de los períodos del modelo de círculos cerrados. Estos modelos en espiral sí son giratorios. Si el eje de rotación es perpendicular a la superficie de la solución se propagan ondas en forma de voluta (como las llamó Winfree) desde el eje, produciendo una estructura en espiral alrededor de una región central que rodea a la parte alta del eje según se mira desde arriba. La estructura es técnicamente una espiral, que se podría dibujar de forma mecánica con un lapicero atado a una cuerda que se va desenrollando de un cilindro central en lugar de hacerlo desde un solo punto. En la espiral química el núcleo central tiene menos de un milímetro de diámetro.

Si la solución es suficientemente profunda (más profunda que el núcleo de la espiral) el eje puede permanecer inclinado hacia un lado. Desde arriba se verán entonces espirales alargadas rodeando la parte central del modelo. En una solución de profundidad adecuada el eje de la voluta puede doblarse para formar una U con su parte superior en una superficie de la solución y la inferior en la otra. Desde una orientación se verán anillos rodeando un punto central. Si la forma de U huye de la superficie que tiene los anillos, éstos desaparecen en esa superficie justo antes de convertirse en círculos completos. Cuando el eje de rotación resulta paralelo a una superficie y en un anillo cerrado, la superficie tiene círculos concéntricos de los que el más interior se propaga hacia dentro y el más externo se propaga hacia fuera.

Si quisiéramos poner barreras al paso de la onda, sería mejor usar una solución más viscosa. Esta se puede conseguir añadiendo dos mililitros de dióxido de silicio coloidal por cada mililitro de solución. Winfree afirma que la solución resultante tiene una consistencia similar a la mantequilla de cacahuete.

Se pueden proteger los modelos y sacarlos de la solución poniendo filtros en la misma. La formación y propagación de las ondas no se alteran por el uso de filtros. Hay que evitar la contaminación de la solución al poner dentro los filtros; para ello deben usarse pinzas de nylon y no se deben tocar los filtros con los dedos. Si se calientan primero los filtros (incluso con el calor del cuerpo) los modelos que se producen son más activos.

Una vez que la solución ha sido absorbida por un filtro, Winfree quita el filtro y lo pone entre dos hojas de plástico (sirven las de envolver comestibles), lo coloca después en aceite para reducir la evaporación y desechar el oxígeno o lo pega a la parte interior de la tapa de un plato de plástico. El filtro se puede usar de nuevo si se aclara bien con agua.

Cuando Winfree introduce un filtro en aceite, puede hacer un juego con las ondas químicas tocando levemente el filtro a través de la película de aceite con un trozo de hierro, como, por ejemplo, el borde de una cuchilla de afeitar. La capa roja resultante, que denomina "la cortina de hierro", bloquea las ondas en esa región durante varios segundos. El tono rojo resulta, aparentemente, del hecho de que el ácido de la solución extrae hierro en estado ferroso de la cuchilla.

Quizás el observador desee fotografiar las ondas a medida que se van desarrollando. Se pueden también fijar en un filtro siguiendo un procedimiento diseñado por Winfree (éste no funcionará con un filtro que haya sido puesto en aceite). Cuando las ondas se estén propagando a través del filtro, se debe retirar éste y ponerlo en agua salada saturada muy fría al objeto de paralizar la reacción. Deje el filtro allí durante cinco minutos (hasta que todos los productos químicos se hayan disuelto excepto la fenantrolina de color rojo ferroso), después séquelo e introdúzcalo en una solución de yoduro sódico (un gramo por cada 40 mililitros de solución con agua) y séquese de nuevo. Después se dejará durante 10 minutos en el vapor de yoduro emanante de una botella conteniendo cristales de yoduro. Este procedimiento fija permanentemente el ferroin.

Enjuáguese el filtro en agua destilada, séquese al aire y aclárese haciéndolo flotar en aceite de parafina; habrá que secarlo otra vez y, finalmente, arroparlo entre dos hojas de plástico adhesivo claro. Winfree afirma que los modelos de ondas químicas preservadas de esta forma han durado años con total agudeza. La conservación durante un día más o menos puede conseguirse de una forma más sencilla sumergiendo los filtros en una solución muy fría de ácido perclórico al 3 por ciento, pero la agudeza disminuye algo y el modelo desaparece cuando el filtro se seca completamente.

Si se desarrollan ondas en un paquete de filtros (se necesitan por lo menos tres por paquete para conseguir las estructuras alargadas), se pueden separar entre sí (como hizo Winfree) para ver las estructuras verticales de las ondas. De esta forma se puede seguir el eje de voluta hacia abajo o desdoblar una estructura en forma de *U*. Las burbujas de dióxido de carbono pueden ser un estorbo si separan los filtros, lo cual ocurre a los cinco minutos más o menos.

Un paquete de filtros Millipore llevan a un mejor entendimiento del anillo de voluta horizontal. En este anillo el eje de la voluta es horizontal y se cierra sobre sí mismo para formar un círculo. Coloque un paquete de cinco filtros en la solución y toque el centro del filtro superior con una aguja caliente para iniciar una onda cilíndrica. Una vez que la onda se ha propagado hacia el exterior, a través del filtro superior haga descender otro grupo de cinco filtros sobre el primer paquete, con el nuevo grupo en el estado rojo inicial de la solución. Pronto aparecerá una onda cilíndrica en la parte alta del nuevo grupo y el círculo se romperá en dos partes, una se moverá hacia dentro y la otra hacia fuera. Cuando se coloca el nuevo paquete de filtros en su sitio, la onda de voluta en el paquete inicial entra en el nuevo paquete y puede llegar a aparecer en lo alto del mismo.

Winfree tiene otra técnica que emplea para preparar los filtros en una demostración. Con ese fin deben mezclarse bromomalonato de amonio uno molar con un volumen igual de fenantrolina ferrosa 0,025 molar. Esta mezcla se combinará con un volumen igual de bisulfato de amonio 4 molar. (46 gramos por cada 100 mililitros de solución con agua.) Coloque un filtro Millipore en la solución, levántelo para que escurra el líquido de la superficie y después póngalo en una superficie de plástico para que se seque. Mientras tanto, ponga un filtro de papel Whatman nº 1 en bromato sódico 0,33 molar (cinco gramos por cada 100 mililitros de solución con agua) y después séquelo. Cuando se quieran demostrar las ondas de voluta, debe mojarse el filtro con agua pura, colocarlo en una superficie limpia (en un plato de plástico) y cubrirlo con el filtro Millipore. Después de esperar unos minutos se verán puntos azules en el filtro de arriba; quizá se vean también ondas azules propagándose a través del filtro a una velocidad de varios milímetros por minuto.

Por 4 dólares puede comprarse una muestra de un filtro Millipore de unos 30 centímetros ya impregnada con los productos químicos para que genere ondas en Winfree (Arthur T. Winfree, Institute for Natural Philosophy, 51 Knoll Crest Court, West Lafayette, Ind. 47906). Se debe enviar una etiqueta con la propia dirección al hacer el pedido. El filtro sólo

necesita ser mojado para que pueda trabajarse con él 30 minutos. El filtro se puede calentar, enfriar, cortar con tijeras, tocar con un objeto de hierro para bloquear las ondas, estimular con una aguja caliente y apilarlo para hacer ondas tridimensionales.

¿Cómo puede una solución mezclada uniformemente a temperatura y presión constantes oscilar de color, de concentración de especies moleculares, de conductividad o de cualquier otra cosa? Esta pregunta ha venido preocupando a los químicos desde el siglo XIX. Las oscilaciones son todavía debatidas y algunos químicos siguen diciendo que las oscilaciones en tales circunstancias son imposibles. En términos científicos la pregunta es, ¿pueden producirse oscilaciones en un sistema homogéneo y cerrado a temperatura y presión constantes? Se sabe que las oscilaciones pueden producirse en medios que tengan constituyentes no homogéneos y gradientes de densidad en las especies moleculares que están sufriendo difusión. Se sabe asimismo que las oscilaciones pueden aparecer si el sistema está abierto a la adición de más materiales. Sin embargo, en las oscilaciones de masa, las oscilaciones se producen, al parecer, en un sistema que es homogéneo y cerrado.

Para desvelar el problema algunos investigadores se concentran en la energía libre de Gibbs del sistema, porque cualquier reacción espontánea en el sistema homogéneo y cerrado a temperatura y presión constantes debe hacer disminuir la energía libre de Gibbs en el sistema. La energía libre de Gibbs es un indicador de la espontaneidad de una reacción (la tendencia de la reacción a continuar por sí misma sin activación externa). En cualquier reacción espontánea, el sistema tiende a reducir su energía (medida por la entalpía) y a incrementar su desorden (medida por la entropía). No es inevitable que ambos extremos puedan ser alcanzados. Por ejemplo, algunas reacciones pueden ser capaces de continuar espontáneamente aun cuando su energía (entalpía) se incremente; la razón es porque el desorden (entropía) se incrementa aún más. Otras reacciones pueden ser espontáneas por la razón opuesta. Así pues, ni la entalpía ni la entropía son buenos indicadores de espontaneidad por sí mismos. Lo que se usa es una combinación de las dos, llamada la energía libre de Gibbs. Cualquier reacción que reduzca la energía libre de Gibbs es aquella que se continúa espontáneamente.

En principio las reacciones de oscilación que he descrito parecen violar la regla universalmente aceptada de que cualquier reacción espontánea debe moverse firmemente hacia su estado del equilibrio final mientras va perdiendo su energía libre de Gibbs. De hecho, las oscilaciones parecen violar la segunda ley de la termodinámica, que prohíbe las oscilaciones en torno al punto de equilibrio. No obstante, la paradoja se resuelve muy fácilmente observando que no hay ninguna ley termodinámica que impida las oscilaciones cuando los reactivos están lejos de equilibrio y a medida que las reacciones netas hacen decrecer la energía libre de Gibbs. Así, en la mezcla de Winfree las reacciones químicas netas consumen muy poco bromato y ácido malónico en cada ciclo de oscilación y los ciclos con los reactivos muy alejados de sus valores de equilibrio finales. En el constante proceso de las reacciones netas hacia el equilibrio, las concentraciones de los componentes menores bromuro y fenantrolina pueden variar considerablemente, aunque permanecen bajas con relación a las concentraciones de bromato y ácido malónico. Con el tiempo, el bromato y el ácido malónico se consumen en las reacciones netas lo suficiente para que el proceso total se pare y todas las concentraciones estén en equilibrio. Así pues, la solución puede oscilar en la concentración de sus especies (con cambio de color en el caso que nos ocupa) siempre que los principales reactivos estén en abundancia y lejos de las concentraciones finales de equilibrio. Una vez que se ha alcanzado el estado de equilibrio cesarán las oscilaciones. No se viola pues la segunda ley de la termodinámica.

Deben existir otros equisitos para que una solución sea capaz de oscilar; de lo contrario, una clase de química sería mucho más colorista. Las reacciones que tienen lugar deben ser complejas, y al menos una debe catalizarse a sí misma (o sea, ser autocatalítica), y enlazarse con las otras reacciones de alguna forma tal como una dependencia de las concentraciones en las otras reacciones o por una inhibición o activación de las reacciones. Este enlace con otras reacciones es común llamarlo realimentación (feedback) y es muy característico de los sistemas de oscilación.

En la mezcla de Winfree el acoplamiento se hace por medio del ion bromuro y la fenantrolina. El bromuro reacciona con el ácido malónico, inhibe la reacción de bromato con la fenantrolina ferrosa y es un producto de la reacción de bromomalonato y fenantrolina férrica. La reacción del bromato con ácido malónico es una reacción autocatalítica. No estoy seguro de que en otros ejemplos de osciladores básicos, como el reloj de yodo por ejemplo, los complicados mecanismos causantes de las oscilaciones hayan sido todavía resueltos y entendidos en términos de termodinámica.

Libros

De la heurística básica del matemático al tratamiento de fuentes bibliográficas

Javier de Lorenzo y José Simón Díaz

PRUEBAS Y REFUTACIONES. LA LOGICA DEL DESCUBRIMIENTO MATEMATICO, por Imre Lakatos. Alianza Universidad. Madrid, 1978. 198 pp. Versión de Carlos Solís. 1. Comprende, esta obra, dos partes: *a) Pruebas y refutaciones (P.R.)* pp. 16-126; *b)* Parte II de *P.R.* y dos Apéndices. La parte *a)* fue publicada por Lakatos en cuatro entregas en *The British Journal for the Philosophy of Science* en 1963-4. La parte *b)* ha sido agregada por Worral y Zahar como homenaje a Lakatos. Tanto una como otra parte proceden de la tesis doctoral de 1961, inédita, de Lakatos, fallecido en 1974. La parte *a)* fue revisada por el autor cuando su publicación, lo que no ocurre con *b)*, donde los editores Worral y Zahar pretenden haber mantenido el contenido aunque no la forma. Ambos editores agregan notas a pie de página indicando dónde creen que Lakatos había cambiado de pensamiento en los años transcurridos desde la elaboración de la tesis y habría modificado el contenido de *P.R.* y, sin duda, de la parte *b).*

2. Aunque mucho se menciona el término "matemática" el desconocimiento que se tiene de esta forma de trabajo es casi total. Desconocimiento también entre quienes se dedican, profesionalmente, al filosofar, al pensamiento. Desconocimiento que, por contraste, viste de unas atribuciones míticas a la Matemática quizá porque se adora, o aborrece, cuanto se desconoce. Y en este desconocimiento se predican de la Matemática múltiples atributos que, en general, no le afectan. Uno de los mitos más comunes es el de la exactitud, infalibilidad, certeza y eternidad del contenido matemático. Otro mito, el de que su método es el estrictamente deductivo, acompañado de que el razonar demostrativo es el modelo del rigor. Mitos que tienen, por supuesto, amplias repercusiones sociológicas –así, en la enseñanza–. Por supuesto, mitos. Desde Galois y Abel hasta Poincaré y Hilbert, los matemáticos creadores han sostenido precisamente lo contrario de tales mitos, afirmando que el hacer matemático depende de la imaginación y que, por ello, jamás puede reducirse a esquema lógico más o menos preestablecido –en todo caso, tal esquema acabará siendo absorbido por el propio hacer matemático–. A pesar de estas opiniones, algunos filósofos mantienen la noción "popular" en torno a la Matemática. Intentando defenderla de su absorción por el logicismo, Poincaré realiza una distinción entre el papel descubridor de la intuición y el plano de la exposición o lógico. Es distinción "redescubierta" por Popper, atribuida a Reichenbach y adoptada por el neopositivismo. Desde esta posición, se rechaza el plano del descubrimiento o matemática ingenua, como plano perteneciente a la psicología y, al adoptar por modo exclusivo el plano de expresión, la Matemática se muestra como un hacer estrictamente formal, reducible a la lógica y de método deductivo puro. En terreno de nadie queda el hacer matemático informal o ingenuo, el que todo matemático hace, de hecho. No formal, para posiciones como el neopositivismo, carece de sentido y de significado; no falsable, por no empírica, resulta que es un hacer que no es tal hacer... Hacer matemático informal o ingenuo, convertido en piedra de toque y contraste de algunas posiciones filosóficas determinadas.

En paralelo a los mitos de las matemáticas, las ciencias han tenido los suyos. Así, el carácter de inductivas que les atribuyó la filosofía baconiana. Inductivismo ingenuo que jamás se ha llevado a efecto en las ciencias –el problema del inductivismo lógico sí es problema, pero nada tiene que ver con el inductivismo experimental al que aquí me refiero–. Como reacción a estos mitos, algunos teóricos de las ciencias han puesto de relieve, precisamente, la ingenuidad de este tipo de inductivismo; entre ellos, Popper, quien ha destacado el papel que el marco previo posee, indicando que de la experiencia desnuda no se obtiene, no se destila ciencia. Siguiendo a Popper, aunque en pugna con él, otros teóricos han vuelto a la historia de las ciencias para, a su través, mostrar el papel de los paradigmas, de los patrones de descubrimiento, de las constelaciones y conjeturas, de la "auténtica" metodología o lógica del descubrimiento... Un acudir a la historia en busca de, como no puede ser de otra manera.

En general, la astronomía, las ciencias físicas han sido los terrenos más solicitados para la mostración de estas tesis y la eliminación del inductivismo. En paralelo, y frente a la tesis del deductivismo matemático, se planteaba la necesidad de invertir dicha tesis, sosteniendo que existe cierto tipo de inducción. Los matemáticos de tendencia constructiva ya venían sosteniendo esta posición, aunque el término "inducción" no puede tener el mismo campo denotativo que en los dominios de las ciencias experimentales. A su pesar, el infalibilismo dogmático se ha mantenido presente en algunos terrenos de la filosofía y de las teorías de las ciencias en relación con las matemáticas. Los mitos son difíciles de erradicar.

3. Es en este contexto en el que hay que situar la obra de Lakatos *Pruebas y refutaciones.* En ella Lakatos pretende adoptar la misma postura crítica que adoptó Popper respecto a las ciencias, pero ahora frente a la Matemática, siguiendo la línea de Polya respecto al papel de la conjetura y contraste en la resolución de problemas. Obra de tesis previa, dogmática, su objetivo se centra en un reto: poner en entredicho alguna de las posiciones del formalismo matemático, de los dogmas del positivismo lógico; aunque se reconozca que no afectará las posiciones últimas del dogmatismo matemático. Desde el comienzo debo precisar: Lakatos identifica formalismo matemático con la posición del neopositivismo lógico que es más bien logicista, no formalista. Su reto se dirige más bien al logicismo que al formalismo, aunque mezcle los términos; reto que por ello cabría suprimir, tras los trabajos de Gödel de 1930. Reto a un mito, por consiguiente, a una imagen "popular" aunque no por ello menos interesante y necesaria. Como reto, se extreman las posiciones y así Lakatos pasa a rechazar todo tipo de deductivismo sin distinguir la variedad de matices que bajo este término se encierran.

La tesis de Lakatos, en cuanto a la metodología o lógica del descubrimiento

matemático, puede sintetizarse en que la matemática informal o ingenua no se desarrolla mediante un aumento de teoremas bien establecidos, sino mediante una continua mejora de las conjeturas gracias a la especulación y a la crítica, apoyadas en una lógica o metodología que califica de "pruebas y refutaciones". Y esta metodología del descubrimiento matemático no es la deductiva, la que va del teorema a su demostración mediante unas reglas de derivación previamente establecidas, sino que responde al esquema metodológico siguiente –y que resume en p. 68 y en 149-50–.

En el trabajo matemático se parte de un problema para cuya resolución se utilizan como reglas heurísticas las siguientes:

1. Se realiza una conjetura primitiva y se contrasta en la mayor cantidad de casos particulares –regla propuesta por Polya–.

2. Hay que intentar probar la conjetura; en la prueba generalmente se habrán introducido lemas implícitos que hay que descubrir en fase calificable de "Análisis de la prueba". El término técnico "prueba" mantiene, en Lakatos, el sentido clásico de experimento mental que sugiere la descomposición de la conjetura primitiva en subconjeturas o lemas.

3. Inmediatamente hay que refutar la conjetura – proceso que Lakatos estima su mayor originalidad–; para ello se buscan contraejemplos bien a los lemas –locales–, bien a la conjetura –globales–.

4. Los contraejemplos o refutaciones locales manifiestan algún lema "oculto" que ahora se explicita y se incorpora como una condición más a la conjetura primitiva. Así se obtiene el teorema, que es la conjetura mejorada con el nuevo concepto generado por la prueba. Si el contraejemplo es global entonces lo que debe rechazarse es la conjetura y lo que hay que hacer es agregar al análisis de la prueba el lema que surge del contraejemplo, lo que conduce a una nueva conjetura en la cual este último lema aparece como una condición más.

Las cuatro reglas heurísticas anteriores constituyen, para Lakatos, el núcleo básico de la lógica del descubrimiento matemático. A ellas pueden agregarse otras reglas, básicamente las de comparar las pruebas con otros teoremas para contraste de la potencia del nuevo lema; comprobar las consecuencias aceptadas de la conjetura primitiva ya refutada; considerar que los contraejemplos se convierten en nuevos ejemplos con lo cual se abren nuevos campos de investigación...

Lakatos, a través de la obra, pretende exponer diversas formas de refutaciones. En particular quiero destacar la calificable como de extensión de conceptos, en el sentido de poner cierto énfasis en la afirmación de que los conceptos no son inmutables, sino que varían: la eternidad de la matemática se viene abajo. Lakatos esboza esta idea en el caso particular de su cuadro metodológico. Al plantear el problema y conjetura consiguiente, se establece la definición de un objeto o se da por establecida; al surgir el contraejemplo, la definición ha de ampliarse para que el lema contenido en el contraejemplo quepa en la conjetura ampliada. Naturalmente, puede rechazarse esta solución y, en lugar de ampliar el concepto cabe restringir la definición primitiva y considerar los contraejemplos como "monstruos" –es lo ocurrido en la teoría de funciones con las funciones "teratológicas" de fines del siglo pasado–. Esta última toma de posición cabe calificarla como de exclusión de monstruos y pertenece, en la terminología de Lakatos, a la heurística lógica. Sin embargo, la posición más aceptable para Lakatos consiste en aceptar tales "monstruos" y ampliar la definición original –es lo que, de hecho, hace el matemático–. Es lo que Lakatos denomina heurística de "contenido". De esta forma, el Análisis de la prueba permite establecer una generalización del primitivo concepto establecido bien en el problema, bien en la conjetura. Con sus palabras: "Las conjeturas ingenuas y los conceptos ingenuos se ven superados por conjeturas (teoremas) mejoradas y por conceptos (generados por la prueba o conceptos teóricos) que se desarrollan a partir del método de pruebas y refutaciones" (p. 111). Es una forma de rechazar la idea de formación autónoma de conceptos así como la afirmación de que la clarificación o explicación de conceptos sea algo preliminar a cualquier discusión científica. Para Lakatos el concepto surge, en matemáticas, a lo largo del proceso de descubrimiento y a lo largo de dicho proceso se transforma y cambia y varía.

Este proceso de ampliación de conceptos plantea, al menos, dos problemas inmediatos que Lakatos no resuelve, esperando en su bosquejo volver quizá a ellos en trabajos posteriores –no conozco que Lakatos haya realizado este tipo de vuelta–. Por un lado, la generalización puede no ser una generalización en todos los casos, sino una auténtica transformación contextual –en su ejemplo es lo que ocurre entre la conjetura de Euler y la demostración topológico-algebraica de Poincaré–; ello obligaría a precisar en qué consiste una y otra, aunque Lakatos parece establecer como criterio que los cambios conceptuales se hacen muy paulatinamente, en alguna de las componentes del concepto primitivo. Por otro, si se tiene un proceso de generalización continuada jamás habría límite alguno en el mismo; sin embargo, de hecho, hay problemas que acaban resolviéndose –bien por solución, bien por disolución–; nada se resuelve acudiendo, como hace Lakatos, a la invariancia de las componentes "lógicas" de las conjeturas, cuando tales componentes no han sido clarificadas. Quizás un trabajo posterior, que Lakatos ya no podrá hacer, hubiera podido aclarar alguna de estas cuestiones, fundamentales para su lógica del descubrimiento.

4. Debo señalar un punto para el lector: el estilo y la forma en que *Pruebas y refutaciones* aparece al público. Contra la expresión dogmática, Lakatos evita el ensayo clásico. Y *P.R.* aparece como un diálogo retomando el estilo del constructivista Heyting en su *Intuicionismo*. El diálogo de Lakatos es entre el Maestro y algunos discípulos muy avanzados, Alfa, Beta, Delta..., en una clase de Matemáticas. El tema sobre el cual se articula el diálogo es el problema de si existe alguna relación entre las caras, vértices y aristas de un poliedro. La conjetura primitiva, la establecida por Euler. Cada uno de los discípulos va expresando en el diálogo las tesis de los diversos matemáticos que se dedicaron al tema desde el siglo XVIII, Cauchy, Lhuillier, Poinsot, hasta llegar a Poincaré quien dio la demostración hoy clásica, pero ya para triangulaciones en símplex y en los terrenos creados por Poincaré de la topología combinatoria, algebraica.

Constituye, de esta forma, y a la vez que una obra de tesis dogmática en cuanto al método utilizado por los matemáticos en su hacer informal, una obra de tesis respecto a la historia. Una historia racional, no real. Y lo que Lakatos denomina aquí historia racional no es otra cosa que una reconstrucción, según su tesis metodológica previa, de lo que en cada instante se produjo respecto al tema elegido. La historia real sólo viene como acompañamiento de la racional y, en el texto, para rematar tal distinción, sólo en las notas a pie de página se recogen las anotaciones históricas "reales". La relación entre ambas historias la ve Lakatos, también en nota: "la historia real es frecuentemente una caricatura de sus reconstrucciones racionales" (p.104).

5. Una frase como la anterior no resuelve problema alguno sino que los plantea, y muy agudos. Son muchos a los que da origen. Si Lakatos se hubiera limitado a desvelar un método, el libro es de gran valor. De hecho, es un esquema, el de las reglas heurísticas, que la exposición de cualquier tesis doctoral de cualquier facultad de matemáticas actual sigue. Natural-

mente, cuando dicha tesis trata de una cuestión –no ya de un problema– particular, interno al contexto de una teoría también determinada y concreta. Y el cierre semántico de esta teoría es un problema que la heurística de Lakatos sería impotente de resolver en su delimitación. Pero Lakatos ha querido ir más allá en la plasmación de un método de trabajo particular: ha querido que este método se prolongue a lo largo de la historia real, condicionando a la propia historia real. Y así ha elegido un tema en el que parece que tal metodología se aplica. Ya en el momento de su publicación se indicó que tal problema era un caso de reconstrucción racional excesivamente particular y "sensible". Los editores han querido subsanar esta crítica agregando los apéndices, aun reconociendo que son incompletos y que Lakatos jamás los completó. En ellos se ponen otros ejemplos, como el de continuidad uniforme, el de medida de Caratheodory..., que no vale la pena discutir al ir en esquema, incompletos e incompletables por el propio autor. Ahora bien, en ellos se resuelve la crítica, porque la crítica sería tangencial. Lo importante, para Lakatos, es no la particularidad sino que este método o lógica del descubrimiento es *el* método matemático auténtico y permanente, aunque los matemáticos ni lo vean, ni lo sigan. Y este es un punto en el que una simple mirada a la historia real permite su falsación, pero negativa para la tesis de Lakatos. Acudir a una historia "racional" no real, lo único que hace es manifestar el carácter ahistórico, precisamente, de la posición de Lakatos.

Y esta ausencia de carácter histórico se muestra en la impotencia de la tesis: su incapacidad de explicar por qué surge este método en los entornos de 1849 y se convierte en el auténtico método del hacer matemático –lo anterior no debe estimarse, pues, matemática–. En su afán antideductivo Lakatos reprocha –un historiador ¿puede reprochar algo a sus personajes?– a Cauchy, a Abel no haber visto dicho método quedándose en el de exclusión de monstruos y delimitación de dominios de validez, argumentando que estaban imbuidos de deductivismo. No ve que son ellos lo que inician un nuevo tipo de hacer matemático en el que la falsación no viene dada por la naturaleza física, por el inductivismo físico, contra el que luchan, sino que ha de crearse y apoyarse en la pura "razón". Son esos matemáticos algunos de los que inician la creación de un nuevo tipo de hacer frente a la matemática científico-natural de Fourier, Laplace... En esta estimación de cambio de cuadro epistemológico y ontológico han de crear nuevos criterios de falsación interna al hacer matemático. Y el primer paso estriba, precisamente, en la determinación de los dominios de validez de los conceptos para, posteriormente, pasar al plano de la refutación. Y el único criterio en este campo es la razón y, con ella, la deducción. Criterio que obligará al establecimiento de una nueva metodología, que es la que Lakatos ha querido poner de relieve. Por supuesto, no independiente, no única respecto a otro tipo de metodologías. Menos aún marginada al contexto de una teoría previa que, para serlo, ha de requerir del cierre semántico y sintáctico, cierres que sólo el enfoque deductivista puede establecer.

Todo ello exige un mayor estudio de los conceptos que se articulan en el problema de la historia. Y el propio Lakatos se hizo consciente de la necesidad de tal clarificación y tras la publicación de *Pruebas y refutaciones*, prestó atención preferente al problema base conceptual de sus reconstrucciones racionales de la historia. Ensayos que no corresponde estudiar aquí, pero que el lector interesado debe conocer –incluso han sido ya traducidos al castellano– para situar en su perspectiva adecuada esta primeriza obra, cuyo valor principal, desde mi punto de vista, reside en haber puesto de manifiesto uno de los métodos de trabajo, una de las heurísticas básicas del trabajador matemático, además del valor referencial de las notas que articulan la historia real en torno a la conjetura de Euler. (J.L.)

TECNICAS DOCUMENTALES Y FUENTES DE INFORMACION, por Nuria Amat Noguera. Bibliograf. Barcelona. 1978. El científico de nuestro tiempo se halla en la imposibilidad de abarcar la información necesaria, que crece a ritmo acelerado, con los medios tradicionales y tiene puesta toda su esperanza en el desarrollo y perfeccionamiento de los sistemas documentales, cuya rápida expansión ha motivado en numerosos países gran variedad de manuales, repertorios y monografías.

En España, la Documentación ha penetrado de dos formas muy diferentes. Por un lado, varias entidades científicas, técnicas y económicas han procurado y conseguido asimilar los conocimientos precisos para aplicarlos a sus tareas y vienen haciéndolo con provecho y eficacia desde hace unos cuantos lustros, mientras que, por otra parte, algunos han estimado que con variar el rótulo de entidades, profesiones o libros de mayor o menor parentesco con la nueva ciencia, era suficiente, de lo cual se ha derivado el recelo con que desde determinados sectores, especialmente el humanístico, viene siendo acogida.

Por todo ello, resultaba indispensable clarificar el panorama, demostrando a la vez la trascendencia de unos métodos indispensables y la falsa idea de que todo se reduce a la aparición de unas máquinas maravillosas que hacen innecesario todo esfuerzo humano, que en último término, según la humorística enumeración del profesor norteamericano que recopiló las concordancias de las obras de Garcilaso de la Vega sirviéndose de diversos tipos de ordenadores, el especialista tendrá que ser la primera y la última de las "máquinas" utilizadas.

Este necesario servicio lo presta, sin duda, el libro de Nuria Amat, por lo que su salida constituye un grato acontecimiento. Sin necesidad de extenderse por terrenos de afinidad dudosa ni en disquisiciones inútiles, muestra la enorme complejidad de la nueva Ciencia y sus infinitas posibilidades, de una manera sistemática, precisa y clara, muy apta para la aplicación didáctica del libro, que, por tal causa, ha de facilitar la enseñanza de una materia cuya introducción en nuevos campos parece inminente. La gran abundancia de esquemas y de gráficos favorece además la comprensión de la doctrina expuesta en párrafos densos y breves.

Es posible que, desde el ángulo del "documentalismo" puro, se discrepe del planteamiento inicial de la autora, en cuanto ésta tiende a subrayar la vinculación que ha tenido desde sus orígenes y debe seguir manteniendo, con cuanto se refiere al libro y por consiguiente con las bibliotecas, pero tan utópico es pretender prescindir de la letra impresa como del hombre, según se indicó anteriormente.

En la primera parte, de carácter introductorio, se estudian el papel de las técnicas documentales en las Ciencias de la Información, las funciones sociales de la documentación y formación documental y las operaciones documentales. El lector no habituado a la materia observará pronto el uso de determinados neologismos, ya admitidos sin discusión, que no son más que el tributo a la nula contribución hispánica al nacimiento y desarrollo de esta Ciencia y al poco éxito de las repetidas tentativas efectuadas durante los últimos decenios para controlar el vocabulario científico y técnico, entre las que destaca el Coloquio celebrado en abril de 1976 por la Real Academia de Ciencias Exactas, Físicas y Naturales con representantes de las similares hispanoamericanas. Quizá resulte más difícil que habituarse al empleo de voces extranjeras, casi siempre inglesas, comprender las nuevas acepciones de otras españolas, como "perfil" o derivados del tipo "utilizador" o "indización" o "repertoriar". Insistimos en que nada de esto debe achacarse a la autora, pues puede hallarse

en cuanto anteriormente se ha escrito en nuestro idioma sobre el tema.

Al tratar de la formación de documentalistas en España y el extranjero, se enúmeran (en la página 28) los organismos que aquí la imparten e incluso los Seminarios, Coloquios y Jornadas que se ocupan de ella, no incluyéndose ninguna referencia a la Facultad de Ciencias de la Información de la Universidad Complutense de Madrid, primer centro de enseñanza superior donde se ha establecido la Documentación como asignatura desde hace algunos años y existe ya un agregado numerario encargado de esta tarea.

La segunda parte trata de las fuentes de información, que define y clasifica meticulosamente, para ocuparse luego de los diversos tipos de centros de información, redes documentales y organismos internacionales. Como en el resto de la obra, cada capítulo lleva una bibliografía complementaria, en su mayoría extranjera, pero en aquellos casos en que existe la nacional se observan algunas omisiones y fallos que convendría subsanar en sucesivas ediciones. La "Selección de fuentes bibliográficas", que ocupa las páginas 92-104, resulta muy vulnerable por la imposibilidad de reducir a esos límites tan dilatada materia. Reducir los "Catálogos de bibliotecas", del mundo, sin ceñirse a los específicos, a siete títulos, de ellos dos españoles; dar algunos otros como "Bibliografías nacionales", junto a repertorios jurídicos, catálogos de librerías y de subastas, boletines de novedades, etcétera, reducir a quince los índices de publicaciones periódicas, dando entre los cuatro españoles los de "Cambio 16" y "Triunfo" y ninguno de los 512 incluidos en la reciente recopilación de Amancio Labandeira Fernández (1976), cuya simple mención habría bastado, etcétera, son algunos de los puntos aludidos. En cuanto a omisiones, cabe destacar la de los diversos trabajos realizados por la mencionada Facultad de Ciencias de la Información y los del Fondo Económico y Social de la Confederación Española de Cajas de Ahorros, así como la de los cuatro volúmenes aparecidos de la "Biblioteca Bibliográfica Hispánica", editada por la Fundación Universitaria Española, de Madrid, que ha empezado a remediar la vergonzosa carencia de una bibliografía de bibliografías españolas. Y respecto a la conveniencia de que donde se ha de inculcar la máxima exactitud y detalle en las referencias bibliográficas, no puedan hallarse ejemplos en contrario, sería menester que desaparecieran algunas citas que tan sólo contienen el autor y el título de la obra.

Para un mayor valor práctico, ha de tenerse presente en todo momento la realidad española, pues lo que en otros lugares es lógico y practicable aquí pertenece al mundo de la utopía. Desde un punto de vista metodológico general, resulta muy convincente el "esquema de criterios de utilización de las fuentes de información bibliográfica" que se ofrece en las páginas 90-92, pero debe pensarse en la perplejidad del "utilizador" de numerosos campos cuando para evitarse la molestia de buscar número a número el contenido de una publicación periódica, opte por dirigirse a los boletines de resúmenes, inexistentes en bastantes materias. Tampoco tendrá demasiada idea del valor de las tesis, al saber que "usualmente tratan de trabajos originales", pues se supone querrá decir que suelen "ser" trabajos originales.

La parte tercera está dedicada al Análisis documental, a través de la descripción bibliográfica (referencia y catalogación), indicación de documentos, resúmenes y lenguajes documentales. También aquí alguna vez se observa el abismo que media entre las grandes aspiraciones teóricas de carácter universal (por ejemplo, sobre los frutos de la aplicación del registro ISBN) y los resultados concretos que sus aplicaciones han tenido entre nosotros.

"La búsqueda documental" constituye la parte cuarta, que informa acerca de los numerosos procedimientos existentes.

La difusión documental y los servicios de difusión (parte quinta) y la automatización documental (parte sexta) representan la utilización práctica de cuanto se ha expuesto anteriormente y combinan los principios doctrinales, las noticias fundamentales sobre los instrumentos que han de manejarse y las normas que han de seguirse, con abundantes ejemplos, aplicables tanto a las necesidades de los investigadores aislados como a las de bibliotecas, centros de documentación, industrias y organismos diversos. La carencia total o la escasez de muchos de los elementos precisos no puede aducirse como argumento para seguir ignorando estos adelantos o impedir que se fomenten las vocaciones hacia actividades tan necesarias y prometedoras. Confiemos en que la formación de los futuros documentalistas españoles se realice al más alto nivel en centros nacidos por el convencimiento de que responden a una exigencia ineludible, no a causa de una moda interesada y efímera, según un plan orgánico que tenga prevista su adecuada utilización posterior. La obra comentada supone una primera notable contribución a esa tarea, ya que ofrece un tratado idóneo para facilitar la comprensión y el conocimiento de una materia que, como se indicaba entre nosotros, se nos venía presentando como algo inasequible o superficial. (J.S.D.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

MONOPOLOS

ON THE DIRAC MAGNETIC POLES. E. Amaldi y N. Cabibbo en *Aspects of Quantum Theory,* dirigido por Abdus Salam y E. P. Wigner. Cambridge University Press. 1972.

HYPOTHETICAL PARTICLES. A. S. Goldhaber y J. Smith en *Reports in Progress in Physics,* vol. 38, pág. 731, 1975.

NEW DEVELOPMENTS IN THE THEORY OF MAGNETIC MONOMOPOLES. P. Goddard y D.I. Olive. Preprint CERN TH. 2445-1977.

LA DIFERENCIACION QUIMICA DE LAS CELULAS NERVIOSAS

CHEMICAL TRANSMISSION BETWEEN RAT SYMPATHETIC NEURONS AND CARDIAC MYOCYTES DEVELOPING IN MICROCULTURES: EVIDENCE FOR CHOLINERGIC, ADRENERGIC, AND DUAL-FUNCTION NEURONS. Edward J. Furshpan, Peter R. MacLeish, Paul H. O'Lague y David D. Potter en *Proceedings of the National Academy of Sciences,* vol. 73, n.° 11, págs. 4225-4229; noviembre, 1976.

RAT SYMPATHETIC NEURONS AND CARDIAC MYOCYTES DEVELOPING IN MICROCULTURES: CORRELATION OF THE FINE STRUCTURE OF ENDINGS WITH NEUROTRANSMITTER FUNCTION IN SINGLE NEURONS. Story C. Landis en *Proceedings of the National Academy of Sciences,* vol. 73, n.° 11, págs. 4220-4224; noviembre, 1976.

NEUROTRANSMITTER SYMPATHETIC AND UPTAKE BY ISOLATED SYMPATHETIC NEURONS IN MICROCULTURES. Louis F. Reichardt y Paul H. Patterson en *Nature,* vol. 270, n.° 5633, págs. 147-151; 10 de noviembre de 1977.

DETERMINATION OF TRANSMITTER FUNCTION BY NEURONAL ACTIVITY Patricia A. Walicke, Robert B. Campenot y Paul H. Patterson en *Proceedings of the National Academy of Sciences,* vol. 74, n.° 12, págs. 5767-5771; diciembre, 1977.

NATURE AND NURTURE IN DEVELOPMENT OF THE AUTONOMIC NEURON. Richard P. Bunge, Mary Johnson y C. David Ross en *Science,* vol. 199, n.° 4336, págs. 1409-1416; 31 de marzo de 1978.

LOS RAYOS COSMICOS DE MAXIMA ENERGIA

COSMIC-RAY SHOWERS. Kenneth Greisen en *Annual Review of Nuclear Science:* vol. 10, dirigido por Emilio Segrè, Gerhart Friedlander y Walter E. Meyerhof. Annual Reviews. Inc., 1960.

COSMIC RAYS. Bruno Rossi. McGraw-Hill Book Company, 1964.

COSMIC RAYS AT GROUND LEVEL. Dirigido por A. W. Wolfendale. The Institute of Physics, Londres, 1973.

IMAGENES NOCTURNAS DE LA TIERRA VISTA DESDE EL ESPACIO

MISSION TO EARTH: LANDSAT VIEWS THE WORLD. Nicholas M. Short, Paul D. Lowman, Jr., Stanley C. Freden y William A. Finch. NASA SP-360. U.S. Government Printing Office, 1976.

APRENDIZAJE Y MEMORIA DE LAS ABEJAS

DAS GEDACHTNIS DER HONIGBIENE FÜR SPEKTRALFARBEN, I: KURZZEITIGES UND LANGZEITIGES BEHALTEN. Randolf Menzel en *Zeitschrift für vergleichenbe Physiologie,* vol. 63, págs. 290-309; 1969.

THE DINAMICS OF LEARNING IN THE HONEYBEE. Jochen Erberen en *Journal of Comparative Physiology,* vol. 99, n.° 3, págs. 231-255; 1975.

NEURAL MECHANISMS OF LEARNING AND MEMORY. Dirigido por Mark R. Rosenzweig y Edward L. Bennett. The MIT Press, 1976.

LA TECNOLOGIA DE LAS MEMBRANAS SINTETICAS

MEMBRANE PROCESSES IN INDUSTRY AND BIOMEDICINE. Dirigido por Milan Bier. Plenum Press, 1971.

POLYMERIC MEMBRANES. R. E. Kesting. McGraw-Hill Book Company, 1971.

INDUSTRIAL PROCESSING WITH MEMBRANES. Dirigido por Robert E. Lacey y Sidney Loeb. Wiley-Interscience, 1972.

REVERSE OSMOSIS MEMBRANE RESEARCH. Dirigido por H. K. Lonsdale y H. E. Podall. Plenum Press, 1972.

EL KIWI

PHYLOGENY AND EVOLUTION OF THE RATITE BIRDS. Joel Cracraft en *Ibis: A Quarterly Journal of Ornithology,* vol. 116, págs. 494-521; 1974.

THE AVIAN EGG: INCUBATION TIME AND WATER LOSS. Hermann Rahn y Amos Ar en *Condor,* vol. 76, n.° 2, págs. 147-152; verano, 1974.

THE KIWI. Brian Reid y Gordon Williams en *Biogeography and Ecology in New Zealand,* dirigido por G. Kuschel. Dr. W. Junk B. V. Publishers, La Haya, 1975.

INCUBATION TEMPERATURES OF THE NORTH ISLAND BROWN KIWI. Barry Rowe en *Notornis,* en prensa.

POKER COMPUTERIZADO

GENERALIZATION LEARNING TECHNIQUES FOR AUTOMATING THE LEARNING OF HEURISTICS. D. A. Waterman en *Artificial Intelligence,* vol. 1 págs. 121-170; primavera, 1970.

STUDIES ON DECISION MAKING USING THE GAME OF POKER. Nicholas V. Findler, Heinz Klein, William Gould, Alexander Kowal y John Mening en *Information Processing 71. Proceedings of IFIP Congress 71, Vol. 2: Applications,* dirigido por C. V. Freiman. North-Holland Publishing Company, 1972.

STUDIES IN MACHINE COGNITION USING THE GAME OF POKER. Nicholas V. Findler en *Communications of the ACM,* vol. 20, págs. 230-245; 1977.

JUEGOS MATEMATICOS

THE CATEGORIES OF CHARLES PEIRCE. Eugene Freeman. Open Court, 1934.

THE PHENOMENOLOGY OF CHARLES S. PEIRCE: FROM THE DOCTRINE OF CATEGORIES TO PHANEROSCOPY. William L. Rosensohn. B. R. Grüner, 1974.

TALLER Y LABORATORIO

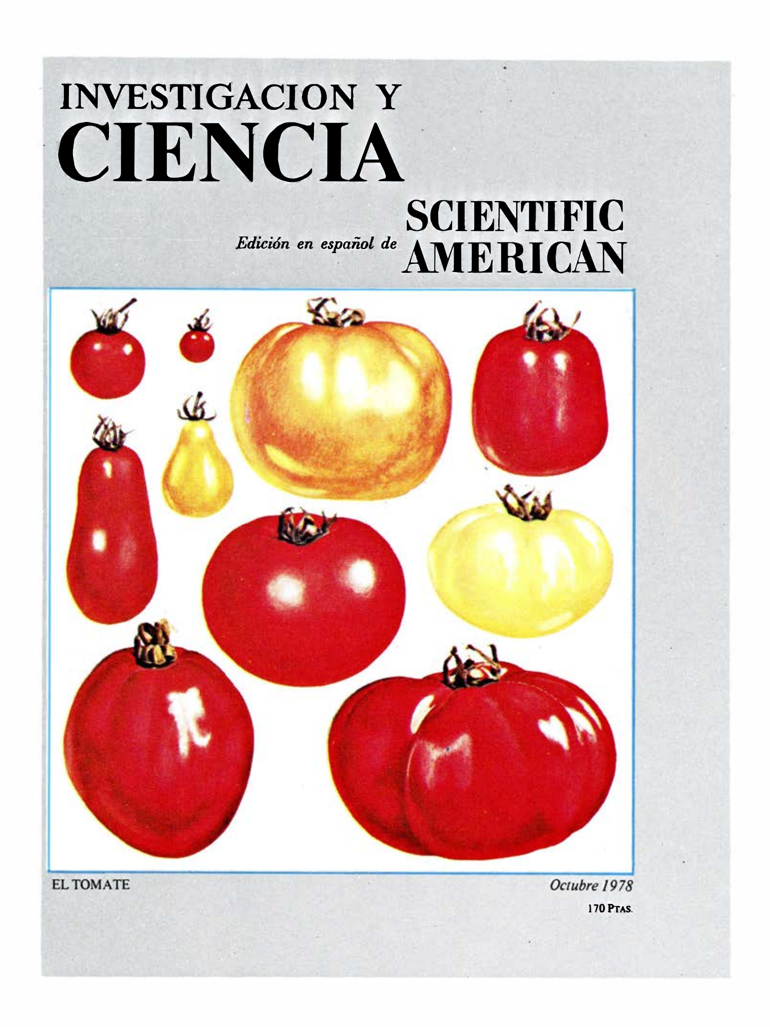
THE COLOR BLIND TRAFFIC LIGHT. John F. Tefelhocz en *Journal of Chemical Education,* vol. 49, n.° 5. págs. 312-314; mayo, 1972.

OSCILLATING CHEMICAL REACTIONS IN HOMOGENEOUS PHASE. Hans Degn en *Journal of Chemical Education,* vol. 49, n.° 5, págs. 302-307; mayo, 1972.

A REACTION PERIODIC IN TIME AND SPACE. Richard J. Field en *Journal of Chemical Education,* vol. 49, n.° 5, págs. 308-311; mayo, 1972.

AN OSCILLATING IODINE CLOCK. Thomas S. Briggs y Warren C. Rauscher en *Journal of Chemical Education,* vol. 50, n.° 7, págs. 496; julio, 1973.

Seguiremos explorando los campos del conocimiento



ORGANISMOS QUE CAPTAN CORRIENTES, Steven Vogel

Gran diversidad de especies, desde las esponjas marinas a los perrillos de las praderas, aprovechan las fuerzas aerodinámicas e hidrodinámicas para aumentar el flujo de aire o de agua que pasa a su través o al de sus moradas.

EL TERMOSTATO DE LOS VERTEBRADOS, H. Craig Heller, Larry I. Crawshaw y Harold T. Hammel

El hipotálamo, estructura situada en la base del cerebro, controla la temperatura corporal en gran variedad de animales y la mantiene a un nivel óptimo mediante la integración de los mecanismos de termorregulación.

EL TOMATE, Charles M. Rick

La especie comestible Lycopersicon esculentum, *considerada durante mucho tiempo como venenosa, se ha convertido en el cultivo alimentario y en la fuente de vitaminas y minerales más extendidos gracias a una amplia modificación genética y a una nueva tecnología de producción.*

COLISIONES DE HADRONES A MUY ALTA ENERGIA, Ramón Pascual y C. Pajares

El comportamiento de las secciones eficaces de hadrones a muy altas energías presenta un carácter crítico similar a las transiciones de fase de segundo orden. Un tal comportamiento puede estudiarse con independencia de los detalles de la cromodinámica cuántica que lo sustenta.

EL PROBLEMA DE LAS TECTITAS, John A. O'Keefe

Estos peculiares trozos de vidrio, ¿se originaron en la tierra o en la luna? ¿Proceden de actividad volcánica o de impactos meteoríticos? Los datos de que se dispone son contradictorios y la solución más razonable es sorprendente.

VIENTOS ESTELARES, Ray J. Weymann

Se observa que estrellas de varias clases y tamaños lanzan materia al espacio formando una corriente lenta y constante. Cierto número de mecanismos pueden ser responsables de este fenómeno sumamente corriente.

LOS HOMINIDOS DE TURKANA ORIENTAL, Alan Walker y Richard E. F. Leakey

Esta región del nordeste de Kenya, a orillas del lago Turkana, encierra el cofre donde se guardan los fósiles de los miembros más antiguos del género Homo.

LA CENTRIFUGA DE GASES, Donald R. Olander

Esta máquina separa los isótopos del uranio mediante centrifugación de un compuesto gaseoso del mismo. Incide eficazmente en el planteamiento de la limitación armamentística.

